



IPTC Standards



Guide for Implementers

Including



Public Release

Document Revision 7.0.

International Press Telecommunications Council
Copyright © 2014. All Rights Reserved
www.iptc.org



This page intentionally blank



Introduction

This document is for all those interested in promoting efficient exchange and re-use of multi-media news content within their own organisations and with information partners, using open standards and technologies.

Whilst a good deal of the content of this document is aimed at technical architects and software writers, business Influencers and decision-makers are encouraged to read the **Executive Summary**, which gives a broadly non-technical justification for the use of IPTC Standards, and how they may be applied to solve the real-world issues of all organisations that create or consume news.

Purpose and Audience

The Guide is intended to provide implementers of the G2 Standards with a thorough knowledge of the XML data structures used to manage and describe content, and an appreciation of the issues involved in implementing the standards in their organisation, whether they are a content provider, content customer, or software vendor.

Terms of Use

Copyright © 2014 IPTC, the International Press Telecommunications Council. All Rights Reserved.

This document is published under the Creative Commons Attribution 4.0 license - see the full license agreement at <http://creativecommons.org/licenses/by/4.0/>. By obtaining, using and/or copying this document, you (the licensee) agree that you have read, understood, and will comply with the terms and conditions of the license.

This project intends to use materials that are either in the public domain or are available by the permission for their respective copyright holders. Permissions of copyright holder will be obtained prior to use of protected material. All materials of this IPTC standard covered by copyright shall be licensable at no charge.

If you do not agree to the Terms of Use you must cease all use of the NewsML-G2 specifications and materials now. If you have any questions about the terms, please contact the managing director of the International Press Telecommunication Council. You may contact the IPTC at www.iptc.org.

While every care has been taken in creating this document, it is not warranted to be error-free, and is subject to change without notice. Check for the latest version of this Document and applicable G2 Standards and Documentation by visiting www.newsml-g2.org and navigating to Documentation resources. The versions of the G2 Standards covered by this document are listed in **About the G2 Standards**.

Contacting the IPTC

IPTC, International Press Telecommunications Council

Web address: <http://www.iptc.org>

Follow us on Twitter: @IPTC

Email: office@iptc.org

Business address

25 Southampton Buildings

London WC2A 1AL

United Kingdom

The company is registered in England at 10 Portland Business Centre, Datchet, Slough, Berks, SL3 9EG

as Comité International des Télécommunications de Presse

Registration No. 1010968, Limited by Guarantee, Not Registered for VAT



Acknowledgments

IPTC member delegates who have contributed to this documentation project (ordered by surname):

BABY, Vincent	Thomson Reuters
BEYNET, Yannick	Agence France Presse
CARD, Tony	BBC Monitoring
COMPTON, Dave	Thomson Reuters
CRAIG-BENNETT, Honor	The Press Association ^{*)}
EVAIN, Jean-Pierre	European Broadcasting Union
EVANS, John	Transtel Communications Ltd
GEBHARD, Andreas	Getty Images
GORTAN, Philipp	APA (Austrian News Agency)
GULIJA, Darko	HINA (Croatia News Agency) ^{*)}
HARMAN, Paul	The Press Association ^{*)}
HUSO, Trond	NTB (Norwegian News Agency)
KELLY, Paul	XML Team
LE MEUR, Laurent	Agence France Presse ^{*)}
LINDGREN, Johan	Tidningarnas Telegrambyrå (Swedish News Agency)
LORENZEN, Jayson	Business Wire
MOUGIN, Philippe	Agence France Presse
MYLES, Stuart	Associated Press
RATHJE, Kalle	Deutsche Presse Agentur
SCHMIDT-NIA, Robert	Deutsche Presse Agentur
SERGEANT, Benoît	European Broadcasting Union
STEIDL, Michael	Managing Director, IPTC
WHESTON, Siobhán	BBC Monitoring
WOLF, Misha	Thomson Reuters

^{*)} Delegate is no longer with the member company in 2014.

Document Author: [Kelvin Holland](#) of Point House Media Ltd.

About the G2 Standards

The Standards covered by this document are:

- ❖ NewsML-G2, version 2.18
- ❖ EventsML-G2, version 2.18
- ❖ SportsML-G2, version 2.1

Document History

Revision	Issue Date	Author (revised by)	Remarks
1	2009-03-20	Kelvin Holland	Public Release
2	2009-11-30	Kelvin Holland	Public Release
3	2011-01-31	Kelvin Holland	Public Release
3.1	2011-02-22	Michael Steidl	Error fixed in 13.4.3
4.0	2011-11-09	Kelvin Holland	Public Release
5.0	2012-11-16	Kelvin Holland	Public Release,
6.0	2014-01-15	Kelvin Holland	Public Release
6.1	2014-03-07	Kelvin Holland	Minor corrections to code samples
7.0	2014-11-07	Kelvin Holland	Public Release

Conventions used by this document

Links to cross-referenced resources within this document are indicated by [this style](#)

Links to external resources are indicated by [this style](#)

Code examples are shown thus:

```
<element attribute="attribute_value">Data</element>
```

All XML elements that consist of two (or more) concatenated words are in lowerCamel case. For example:

```
<catalog>
<catalogRef>
```

Where a word is normally capitalized, it remains so. Thus:

```
<inlineXML>
```

Attribute names are always all lowercase For example

```
standardversion="2.18"
```

The term "Item" with capitalised "I" indicates a NewsML-G2 Item (e.g. News Item, Planning Item, etc.).



The "red hand" indicates especially important notes or warnings to implementers

Note on Spelling (English)

The IPTC convention for documents in English is to use UK English spelling. In general, U.S. English is used for property names and values used in IPTC XML Standards (for example, canceled, color, catalog).

A common sense approach dictates that there may be exceptions to this convention.

Terminology: MUST and SHOULD

There are few mandatory features in NewsML-G2. This document uses the terms MUST (NOT), SHOULD (NOT) and MAY as defined in RFC 2119. A MUST instruction in this document refers to mandatory actions; SHOULD refers to recommended actions or best practice to be used unless there is a very good reason not to do so; MAY refers to optional actions.

Note on Time and Date-Time properties

The XML Specification for time-based properties is based on ISO 8601 and permits the omission of time zone/time offset information. However, these values MUST be present in NewsML-G2 timestamps that express a date AND time, such as Item Metadata timestamps, because the exchange of news information may cross time zones, and timing information must be unambiguous. The following comply:

```
<versionCreated>2014-11-06T12:12:12+01:00</versionCreated>    offset +
<versionCreated>2014-11-06T12:12:12-01:00</versionCreated>    offset -
<versionCreated>2014-11-06T12:12:12Z</versionCreated>         indicates UTC
```

The following does NOT comply:

```
<versionCreated>2014-11-06T12:12:12</versionCreated>         no time offset
```

Administrative content metadata properties such as <contentCreated> and descriptive date-time properties such as <founded> and <dissolved> use a Truncated DateTime data type permitting parts of the data-time to be truncated from the right, for example:

```
<founded>2014-11-06</founded>
```

complies, but this does not:

```
<founded>2014-11-06T12:00:00</founded>
```



This page intentionally blank



Table of Contents

Introduction	3
Purpose and Audience	3
Terms of Use.....	3
Contacting the IPTC	3
Acknowledgments	4
About the G2 Standards.....	4
Document History	4
Conventions used by this document.....	5
Table of Contents	7
Table of Figures.....	13
Table of Code Listings.....	15
1 How-To Index.....	17
1.1 General design	17
1.2 Implement semantic technology concepts.....	17
1.3 Manage all kinds of content.....	17
1.4 In more detail.....	17
1.5 Migrate to G2 from legacy formats	18
1.6 History of changes	18
2 What's New in this release: NewsML-G2 2.18.....	19
2.1 Extending NewsML-G2	19
2.2 Extended metadata about Controlled Vocabularies.....	19
2.3 Time precision improvement for video.....	19
2.4 Enhancements to news coverage	19
2.5 Special processing can be signalled for part of content	19
2.6 Extending the use of full URIs as concept identifiers	19
3 Executive Summary.....	21
3.1 Why News Exchange standards matter.....	21
3.2 The purpose of NewsML-G2	21
3.3 Business Drivers.....	21
3.4 Business Requirements.....	21
3.5 Design and Benefits	22
4 About the IPTC.....	23
5 Unification of NewsML-G2 and EventsML-G2.....	23
6 How News Happens	25
6.1 Introduction.....	25
6.2 Assignment	25
6.3 Information Gathering.....	25
6.4 Verification	26
6.5 Dissemination.....	26



6.6	Archiving.....	27
7	Quick Start: NewsML-G2 Basics	29
7.1	About the Quick Start Guides.....	29
7.2	Introduction.....	29
7.3	Item structure	29
7.4	“Invisible” values	30
7.5	Root element	30
7.6	Item Metadata <itemMeta>	33
7.7	Content Metadata <contentMeta>	35
7.8	Content.....	38
7.9	Summary and Next Steps.....	40
7.10	Hidden Values of NewsML-G2	40
8	Quick Start: Text.....	41
8.1	Introduction.....	41
8.2	Example	41
8.3	Document structure	41
8.4	Text content choices	43
8.5	Putting it together.....	44
9	Quick Start: Pictures and Graphics	45
9.1	Introduction.....	45
9.2	About the example	45
9.3	Document structure	45
9.4	Item Metadata <itemMeta>	46
9.5	Embedded metadata	46
9.6	Content Metadata <contentMeta>	47
9.7	Picture data.....	50
9.8	The Complete Picture	54
10	Quick Start: Video	57
10.1	Introduction	57
10.2	Part I – Multiple Renditions of a Single Video	57
10.3	Audio metadata	61
10.4	Full Part 1 listing.....	62
10.5	Part 2 – Multi-part video	63
10.6	Introduction	63
10.7	Full Part 2 listing.....	65
11	Quick Start - Packages	69
11.1	Introduction	69
11.2	Packages and Links: the difference.....	69
11.3	Document structure.....	70
11.4	Item Metadata	70
11.5	Content Metadata.....	70



11.6	Package Structure.....	71
11.7	Complete Listing of a Simple Package	72
11.8	Hierarchical Package Structure.....	73
11.9	List Type Package Structure	74
11.10	A Sequential “Top Ten” Package.....	76
11.11	Package Processing Considerations	78
12	Concepts and Concept Items	81
12.1	Introduction	81
12.2	What is a Concept?.....	81
12.3	Creating Concepts – the <concept> element.....	81
12.4	Conveying Concepts: the Concept Item structure.....	83
12.5	Concepts for real-world entities.....	84
12.6	More real-world entities.....	88
12.7	Relationships between Concepts.....	90
12.8	Supplementary information about a Concept	92
12.9	Concepts in Practice	93
13	Knowledge Items	95
13.1	Introduction	95
13.2	Example: a Knowledge Item for <accessStatus>	95
13.3	Structure and Properties.....	96
13.4	Knowledge Workflow.....	99
13.5	Using G2 Knowledge Items with SKOS.....	100
14	Controlled Vocabularies and QCodes	103
14.1	Introduction	103
14.2	Business Case	103
14.3	How QCodes work	103
14.4	QCodes and Taxonomies	106
14.5	Managing Controlled Vocabularies as NewsML-G2 Schemes.....	106
14.6	Creating and Managing Catalogs.....	108
14.7	Processing Catalogs and CVs	111
14.8	Private versions and extensions of CVs.....	115
14.9	Best Practice in QCode exchange.....	118
14.10	Syntactic Processing of QCodes	119
14.11	A generic way to express concept identifiers as URIs.....	119
14.12	Literal Identifiers	120
14.13	Full list of the QCode type attributes and their URI siblings	121
15	EventsML-G2	123
15.1	A standard for exchanging news event information.....	123
15.2	Event Information – What, Where, When and Who	125
15.3	Event Coverage.....	128
15.4	Event Properties in more detail.....	128



15.5	Multiple Event Concepts in a Knowledge Item.....	140
15.6	Conveying Events in a NewsML-G2 Package.....	144
15.7	Events Workflow	145
15.8	Events in a News Item	145
16	Editorial Planning – the Planning Item.....	149
16.1	Item Metadata <itemMeta>	149
16.2	Content Metadata <contentMeta>	150
16.3	<newsCoverageSet>	150
16.4	<newsCoverage>	150
16.5	<planning>.....	150
16.6	The <delivery> component.....	156
16.7	Delivered Items - <deliveredItemRef>	156
17	SportsML-G2	159
17.1	Introduction	159
17.2	Business Advantages	159
17.3	Structure	160
18	Exchanging News: News Messages.....	173
18.1	Introduction	173
18.2	Structure	173
19	Migrating IPTC 7901 to NewsML-G2.....	179
19.1	Introduction	179
19.2	Message Header.....	179
19.3	Message Text.....	180
19.4	Post-Text Information	180
20	Designing a NewsML-G2 feed from scratch.....	183
20.1	Introduction	183
20.2	Content-driven design	183
20.3	The Basics	184
20.4	Identification and version control	184
20.5	Conformance	184
20.6	Validation	184
20.7	Catalogs and Controlled Vocabularies.....	184
20.8	Rights Information.....	184
20.9	Management Metadata.....	184
20.10	Property Usage.....	185
20.11	IPTC Controlled Vocabulary Usage	188
21	News Industry Text Format (NITF).....	189
21.1	Introduction	189
21.2	Overview of NITF and NewsML-G2 equivalents.....	190
21.3	Content metadata <body.head>	192
21.4	Body Content <body.content>.....	193



21.5	End of Body <body.end>	193
22	Mapping Embedded Photo Metadata to NewsML-G2	195
22.1	Introduction	195
22.2	IPTC Metadata for XMP	195
22.3	Synchronizing XMP and legacy metadata.....	196
22.4	Picture services using IIM/XMP	196
22.5	Rationale for moving to G2.....	196
22.6	IIM resources	197
22.7	Approach	197
22.8	IIM to NewsML-G2 Field Mapping Reference.....	197
22.9	IIM to NewsML-G2 Mapping Example.....	198
22.10	Reconciling NewsML-G2 with embedded metadata.....	205
23	Rights Metadata.....	207
23.1	Rights Info (Core Conformance Level).....	207
23.2	Link to an external Rights Information resource.....	207
23.3	Business Reasons for Extended Rights Information at PCL	208
23.4	The extended <rightsInfo> model	208
23.5	Use cases for the extended <rightsInfo> model	209
23.6	Processing models for extended Rights Info.....	212
24	Expressing company financial information	213
24.1	Background	213
24.2	<hasInstrument> properties (as attributes)	213
24.3	Adding <hasInstrument> to a G2 item	214
25	Advanced Metadata Techniques	217
25.1	Introduction	217
25.2	The Assert wrapper	217
25.3	Inline Reference.....	221
25.4	Why and How metadata has been added: @why and @how	223
25.5	Adding customer-specific metadata	224
25.6	The derivation of metadata: the <derivedFrom> element.....	226
25.7	Generic properties for proprietary use.....	227
25.8	Special processing instructions for metadata about a part of the content	228
26	Generic Processes and Conventions	229
26.1	Processing Rules for NewsML-G2 Items	229
26.2	Using Links	229
26.3	Publishing Status	232
26.4	Embargo.....	233
26.5	Geographical Location	234
26.6	Processing Updates and Corrections.....	235
26.7	Content Warning	237
26.8	Working with Social Media.....	238



26.9	Indicating that a News Item has specific content	241
27	Identifying Sources and Workflow Actors	243
27.1	Introduction	243
27.2	Information Source <infoSource>	243
27.3	Identifying workflow “actors” – a best practice guide.....	243
27.4	Transaction History <hopHistory>	244
27.5	Hash Value <hash>	246
28	Changes to G2-Standards Family.....	249
28.1	Introduction	249
28.2	Guidelines release 7 – covers NewsML-G2 2.16 --> NewsML-G2 2.18	249
28.3	Guidelines release 6 – covers NewsML-G2 2.13 --> NewsML-G2 2.15	249
28.4	Guidelines release 5 – covered NewsML-G2 2.10 --> NewsML-G2 2.12	249
28.5	Guidelines release 4 – covered NewsML-G2 v2.7 -EventsML-G2 1.6 →NewsML-G2/EventsML-G2 2.9	251
28.6	News Architecture (NAR) v1.7 → v1.8	252
28.7	NewsML-G2 v2.6 → v2.7.....	253
28.8	EventsML-G2 v1.5 → v1.6.....	254
28.9	News Architecture (NAR) v1.6 → v1.7	254
28.10	NewsML-G2 v2.5 → v2.6.....	254
28.11	EventsML-G2 v1.4 → v1.5.....	254
28.12	News Architecture (NAR) v1.5 → v1.6	254
28.13	NewsML-G2 v2.4 → v2.5.....	255
28.14	EventsML-G2 v1.3 → v1.4.....	255
28.15	News Architecture (NAR) v1.4 → v1.5	255
28.16	NewsML-G2 v2.3 → v2.4.....	257
28.17	EventsML-G2 v1.2 → v1.3.....	257
28.18	News Architecture (NAR) 1.3 → 1.4	257
28.19	NewsML-G2 v2.2 → v2.3.....	258
28.20	EventsML-G2 v1.1 → v1.2.....	258
28.21	SportsML-G2 v2.0 → v2.1	258
29	Additional Resources.....	259



Table of Figures

Figure 1: The heart of NewsML-G2 is the News Architecture (NAR); framework that supports many different information needs using common components	22
Figure 2: All G2 Items share this basic container structure	29
Figure 3: The basic XML elements associated with each part of a G2 Item.....	30
Figure 4: IPTC Core Metadata fields in the File Info panel of Adobe Photoshop	47
Figure 5: Each <remoteContent> wrapper references a separate rendition of the binary resource.....	51
Figure 6: A Top Ten News Package displayed on the Web	69
Figure 7: Top-level element view of a simple package, and (right) a visualisation of the structure	71
Figure 8: Code outline of hierarchical package with two groups, visualising parent-child structure (right)	73
Figure 9: Code skeleton view of package with alternative components, with visualisation of structure	74
Figure 10: Code skeleton of a sequential mode package and (right) the resulting relationship structure ..	76
Figure 11: Diagrams of Package Profiles. The numbers in brackets indicate the required items.....	79
Figure 12: Information Flow for Concepts and Knowledge Items	100
Figure 13: Human-readable browser page of a Concept URI.....	105
Figure 14: Persistent Event as a Concept using <eventDetails> and (right) transient Events in a News Item.....	125
Figure 15: Hierarchy of Events created using Event Relationships.....	130
Figure 16: Event Flow using Package Items and Knowledge Items.....	145
Figure 17: Planning Item: the newsCoverageSet has one or more newsCoverage components	150
Figure 18: Structure of a SportsML-G2 document matches NewsML-G2	160
Figure 19: Diagram of Package Group structure.....	170
Figure 20: News Messages can convey any number of any kinds of G2 Item.....	173
Figure 21: The IPTC Core metadata panel for an image opened in Adobe Photoshop.....	195
Figure 22: IPTC fields in Photoshop File Info panel (photo and metadata ©Thomson Reuters).....	198
Figure 23: State Transition Diagram for <pubStatus>	233
Figure 24: Processing <embargoed>.....	234



This page intentionally blank



Table of Code Listings

LISTING 1 A NewsML-G2 News Item	38
LISTING 2 NewsML-G2 Text Document.....	44
LISTING 3 Photo in NewsML-G2.....	54
LISTING 4 Multiple Renditions of a Video in NewsML-G2	62
LISTING 5 Multi-part Video in NewsML-G2.....	65
LISTING 6 Simple NewsML-G2 Package.....	72
LISTING 7 Group Set example showing Hierarchical Package Structure	73
LISTING 8 Group Set example showing an “alt” Package Mode.....	75
LISTING 9 Group Set example showing a “seq” Package Mode.....	76
LISTING 10 Abstract Concept conveyed in a NewsML-G2 Concept Item.....	83
LISTING 11 Person Concept conveyed in a NewsML-G2 Concept Item.....	86
LISTING 12 Knowledge Item for Access Codes.....	98
LISTING 13 Complete Catalog Item	111
LISTING 14 Event sent as a Concept Item.....	127
LISTING 15 Two Related Events in a Knowledge Item.....	141
LISTING 16 A Set of Events carried in a News Item	146
LISTING 17 Planning Item at CCL.....	150
LISTING 18 Planning Item at PCL	155
LISTING 19 Planning Item with delivery at CCL.....	156
LISTING 20 Planning Item with delivery at PCL	157
LISTING 21 A complete sample SportsML-G2 document.....	166
LISTING 22 Sports story in NewsML-G2/NITF	168
LISTING 23 SportsML-G2 Package	171
LISTING 24 News Message conveying a complete News Package	176
LISTING 25 An NITF marked-up article conveyed in <inlineXML>	193
LISTING 26 Embedded photo metadata fields mapped to NewsML-G2	204
LISTING 27 Company Financial Information.....	215
LISTING 28 Illustrating Located, Subject and Dateline	235
LISTING 29 Hop History	245



This page intentionally blank



1 How-To Index

The Implementation Guide is written using worked examples and use cases in order to give implementers an insight into the practical application of NewsML-G2 features. It may be helpful to have the NewsML-G2 Specification at hand; this will contain further detailed information about features discussed. This document is part of the package containing this Guidelines document, which can be downloaded by visiting www.newsml-g2.org and navigating to Specification resources.

Implementers picking up these full Guidelines after reading the separately available “Quick Start” guides may wish to go straight to the How To topics for answers to specific questions.

1.1 General design

[Learn the **shared basics** of all types of a NewsML-G2 Item](#)

[Get advice on designing a **NewsML-G2 feed from scratch**](#)

[View **frequently used** NewsML-G2 properties](#)

[Reveal **hidden values** of NewsML-G2 properties](#)

[Add **proprietary properties** to NewsML-G2](#)

1.2 Implement semantic technology concepts

[Express **structured metadata** about people, organisations, places and objects](#)

[Create and manage **Controlled Vocabularies**](#)

[Processing **QCodes**](#)

1.3 Manage all kinds of content

[Manage **text content** of a News Item](#)

[Manage **picture content** of a News Item](#)

[Manage **video content** of a News Item](#)

[Manage content of a **Package Item**](#)

[Exchange **news** using News Messages](#)

[Create, send and update information about **news events**](#)

[Share **news planning and fulfilment** information with partners and customers](#)

[Convey **sports** information](#)

1.4 In more detail

[Learn about **timestamps** of a NewsML-G2 item](#)

[Apply the **Publishing Status** of a NewsML-G2 Item](#)

[Apply **embargo** to news](#)

[Process **updates**](#)

[Process **corrections**](#)

[Express **warnings** about content](#)

[Apply **Editorial Notes** to a NewsML-G2 Item](#)

[Express **rights metadata**, from simple to fine-grained](#)

[Categorize content of a News Item](#)



[Express financial market information about companies](#)

[Convey quantitative data, such as currency amounts](#)

[Assign social media tags to news](#)

[Add social media ratings to content](#)

[Identify sources and workflow actors](#)

[Create links from content to inline metadata](#)

[Add user-customized metadata to a NewsML-G2 Item](#)

1.5 Migrate to G2 from legacy formats

[Migrate a news feed from IPTC7901 to NewsML-G2](#)

[Migrate embedded photo metadata to NewsML-G2](#)

[Migrate NITF management metadata to NewsML-G2](#)

1.6 History of changes

[View changes to NewsML-G2 Standards from previous versions](#)



2 What's New in this release: NewsML-G2 2.18

This chapter summarises the changes to NewsML-G2 since version 2.15, which was covered in the previous revision of this Guidelines document. A full change history of NewsML-G2 from version 2.2 is documented in [28 Changes to G2-Standards Family](#).

2.1 Extending NewsML-G2

The ability to extend NewsML-G2 to create proprietary properties has been a feature from the outset, but until now this required the use of proprietary properties. Such properties need to be created in a separate namespace and therefore can not be processed by a standard G2 property software module.

To overcome this limitation, a new way to extend NewsML-G2 has been added, using a set of generic G2 properties created for this purpose, one for each G2 “wrapper” element (for example the generic extension element for `<itemMeta>` is `<itemMetaExtProperty>`). The semantics of a generic property are defined by the concept referenced by the property's `@rel` attribute. See [25.7 Generic properties for proprietary use](#).

2.2 Extended metadata about Controlled Vocabularies

When a Knowledge Item is used to deliver the complete set of concepts in a scheme, it is now possible to express metadata for the scheme using `<schemeMeta>`, rather than relying on the use of `<scheme>` in a Catalog.

The new element also includes a `<related>` child element, and this allows the expression of relationships at the scheme level. See [13.3.5 Scheme Metadata](#).

2.3 Time precision improvement for video

Alternative renditions of a video may use different frame rates, and in these circumstances a single value denoting the time delimiter for the different shots of a multipart video will not be frame-accurate. This has been corrected by the addition of `@renditionref` to the `<timeDelim>` element of part metadata. See [10.6.1.2 Timing metadata](#).

2.4 Enhancements to news coverage

The `<newsCoverage>` element has had Extension Points added and the new `<newsContentCharacteristics>` element allows the expression of physical properties of the planned news coverage (for example the height and width of an image) for video and still images. See [16.5.1 <planning> at Power Conformance Level](#).

2.5 Special processing can be signalled for part of content

Providers are now able to specify processing instructions that can be applied to different parts of an Item's content.

As this change highlights the need for implementers to be aware of inheritance of property values by child elements from their parents, the general G2 rule on inheritance as it applies to `<partMeta>` is also re-stated for clarification. See [25.8 Special processing instructions](#).

2.6 Extending the use of full URIs as concept identifiers

When NewsML-G2 was originally designed, communication capacity constraints required a compact way to express the typically long globally unique identifier (URI) of a concept, which led to the creation of the QCode as a method of abbreviating URIs. In G2 version 2.11, `@uri` was introduced for properties having a concept identifier value type, restricted to elements that used `@qcode` or `@literal` to hold the identifier

This has now been extended by introducing a corresponding full URI type alternative for the existing QCodeType and QCodeListType attributes, for example `@role` gets a `@roleuri` sibling. See [14.11 A generic way to express concept identifiers as URI](#).



This page intentionally blank

3 Executive Summary

3.1 Why News Exchange standards matter

Information is valuable. Many major financial decisions rely on split-second delivery of news about companies and markets; successful businesses have been built on the ability to target individuals and groups with information which is relevant to their needs. News organisations and information providers have also invested heavily in the people and the technology needed to gather and disseminate news to their customers.

Without standards for news exchange, most of the value of this information would be lost in a confusion of customised feeds and competing formats. The huge volumes of content now being exchanged not only demand a common format, or mark-up, for the content itself but also a common framework for information about the content - the so-called metadata.

3.2 The purpose of NewsML-G2

NewsML-G2 is an open standard for the exchange of all kinds of news content, be it text, pictures, audio or video. There are also related standards for expressing news events and sports information: EventsML-G2 and SportsML-G2¹. The content can be in any format or encoding. It is conveyed with semantically-rich metadata that matches the needs of a professional news workflow and the way that content is consumed both in a B2B and B2C context.

The standard is not concerned with the presentation or mark-up of the content that it conveys; this is the role of standards such as HTML5 and NITF (News Industry Text Format, an IPTC XML standard) that are used to mark up the payload of a NewsML-G2 Item. Microformats such as hNews are complementary; the IPTC has its own semantic mark-up standard rNews, which is compatible with NewsML-G2.

NewsML-G2 models the way that professional news organisation work, but goes beyond this by standardising the handling of the metadata that ultimately enables all types of content to be linked, searched, and understood by end users. NewsML-G2 metadata properties are designed to be mapped to RDF, the language of the Semantic Web, enabling the development of new applications and opportunities for news organisations in evolving digital markets.

3.3 Business Drivers

Many of the business challenges faced by media organisations are related to the development of the World Wide Web, which has not only increased the availability of news content, but is constantly creating new ways to consume it. These challenges are not a once-in-a-lifetime event, but a continuing fact of life.

Businesses need to:

- ❖ Control, and if possible reduce, the cost of developing and maintaining services.
- ❖ Quickly develop new media-rich products and services that can exploit emerging trends and new business models.
- ❖ Give customers access to added-value assets, including archives and metadata repositories.
- ❖ Allow innovation by third-party vendors and partners.
- ❖ Enhance IT investment by enabling the sharing of complex content across separate systems.

3.4 Business Requirements

In response to these business challenges, an information exchange standard needs to:

- ❖ Fit an MMM strategy (Multi-media, Multi-channel, Multi-platform)
- ❖ Handle texts, pictures, graphics, animated, audio or video news

¹ EventsML-G2 expresses news events, including calendar items and breaking news.

- ❖ SportsML-G2 is a mark-up language for sports results and statistics using a plug-in architecture for recording highly-detailed actions for specific sports (e.g. ice hockey, football, baseball)

- ❖ Be a lightweight container for news, simple to implement and extend, yet offer powerful features for advanced applications.
- ❖ Be useful at all stages of the lifecycle of news, from initial event planning, through content gathering, syndication, to archiving.

3.5 Design and Benefits

The NewsML-G2 standards are based on common framework – the News Architecture – that is independent of any technical implementation. It may be implemented using object-oriented software, such as Java, or in a database. The IPTC has implemented the NAR specification in XML Schema to create NewsML-G2 and its “siblings”, EventsML-G2 and SportsML-G2, because of the need to facilitate news exchange using Internet (W3C) standards. XML provides continuity with existing standards, and also has an existing large community of experts.

■ *Everything should be made as simple as possible, but not one bit simpler*
Albert Einstein

The standard enables all parties involved in news – providers, receivers and software vendors – to send and receive information quickly, accurately, and appropriately.

- ❖ **A common framework** maximises the value of investments and provides a path into the future, with maximum inter-operability between different information partners.
- ❖ **Machine-readable** metadata enables automation of standard processes, cutting costs, speeding delivery, and increasing quality.
- ❖ **Innovative solutions** are possible because NewsML-G2 complements the work of companies working on search and navigation technologies to realise the vision of the Semantic Web.

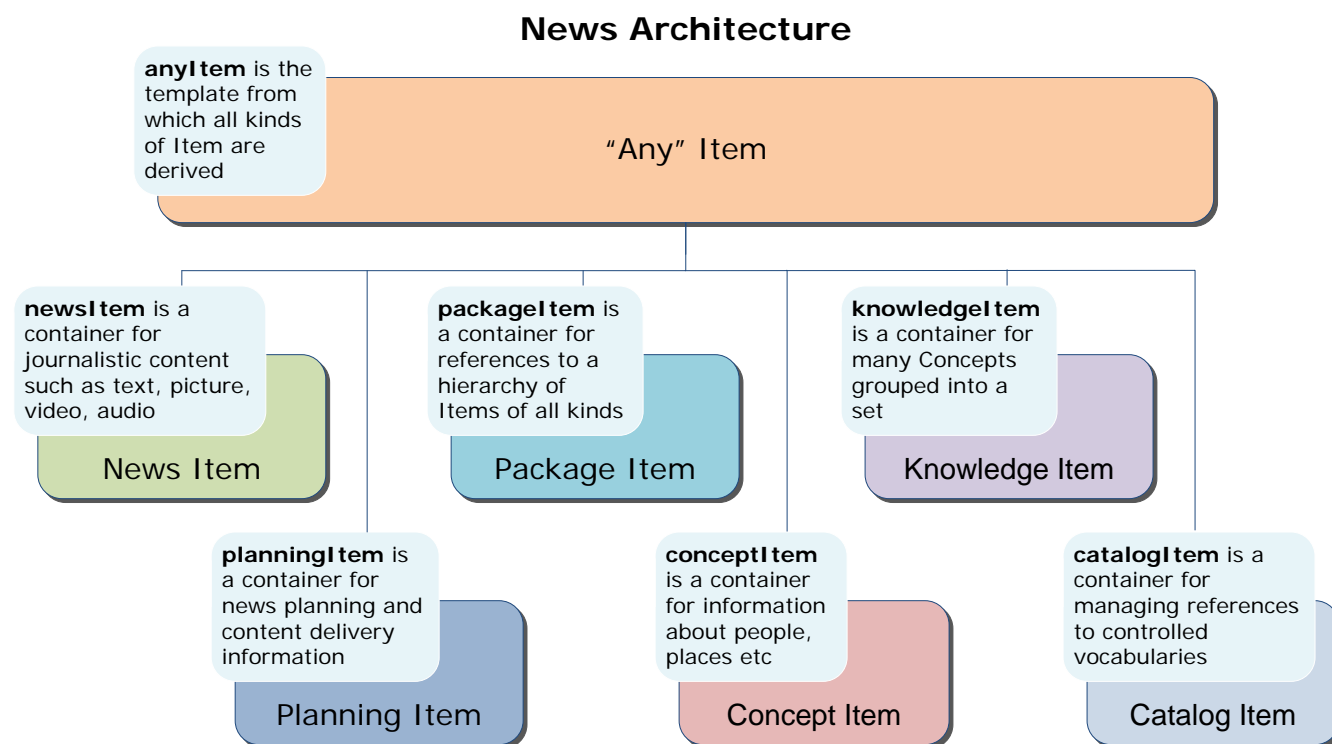


Figure 1: The heart of NewsML-G2 is the News Architecture (NAR); framework that supports many different information needs using common components



4 About the IPTC

IPTC members are technologists and thought leaders from the world's main news agencies and leading media players. They are expert in the field of news production and dissemination. IPTC Standards today play an essential role in efficient news exchange between the world's news and media organisations.

The text transmission standards IPTC7901 and its cousin ANPA1312 have been a key enabler of news exchange for news agencies and their newspaper and broadcaster customers, as has the IIM standard for pictures. All media organisations have benefited from these standards; they have been essential to the adoption of digital production technologies. NewsML was first launched in 2000, and the G2 version of NewsML in 2006. This has since been followed by EventsML-G2 and SportsML-G2.

5 Unification of NewsML-G2 and EventsML-G2

Originally, NewsML-G2 and EventsML-G2 were separate G2 Standards, although they shared a common base. With the introduction of the Planning Item in NewsML-G2 v2.7 and EventsML-G2 v1.6, it became difficult to make a distinction between the two as separate standards.

The IPTC members therefore decided to unify the Standards. From a non-technical "brand" standpoint, NewsML-G2 is now the "senior" umbrella standard for all G2 Items, whether one of its Items conveys purely NewsML-G2 or EventsML-G2. SportsML-G2 continues to be a completely separate standard, although it is always conveyed by NewsML-G2. From v2.9, the name EventsML-G2 can continue to be used for the feature of conveying events, but all of its structures are now merged in NewsML-G2.

From a technical implementation perspective, there is a simplification of XML Schemas into just two: an "All-Core" Schema for Core Conformance, and an "All-Power" Schema for the Power Conformance Level. These cover all Item types in both NewsML-G2 and EventsML-G2.



Throughout this document, the term "NewsML-G2" is used, unless referring to a specific feature of an earlier version of EventsML-G2. The term "G2" on its own refers to the G2 Family of Standards.

The term "Item" with capitalised "I" is used to indicate a NewsML-G2 Item (News Item, Planning Item, Package Item etc.).



This page intentionally blank

6 How News Happens

6.1 Introduction

NewsML-G2 represents a content and processing model for news that aligns with the way that professional news organisations work. It is therefore important that implementers have at least a high-level understanding of how the news business works, in order to appreciate the rationale behind its features.

An event becomes news when someone decides to create a record of it, and place that record in the public domain. Professional news production is not a haphazard or random process, but a highly organised activity, shaped by a number of influences:

- ❖ The publishing of news originally centred on printing, an industrial process which imposes time and logistical constraints. Print remains an important channel for news dissemination.
- ❖ The selection of what is, and what is not, news to any given audience is vital to the success of any publishing venture, whether in print, broadcasting, web or other media.
- ❖ For legal and ethical reasons, professional news organisations ensure that standards are maintained in the selection and production of news, and that content is reviewed before being authorised for release to the public

These constraints and considerations lead to the news production process being divided into five generic domains:

- ❖ Planning and Assignment
- ❖ Information Gathering
- ❖ Verification
- ❖ Dissemination
- ❖ Archiving

6.2 Assignment

News organisations need to plan their operations, based on prior knowledge of newsworthy events that are expected to occur in any given time frame (daily, weekly, monthly etc). The resulting schedule of events is called a variety of names, according to custom, such as schedule, budget, day book or diary.

Unexpected events (breaking news) will cause this schedule to change at short notice.

According to the schedule, people and resources will be assigned to “cover” the news events, and those who are dependent on the timely gathering of the news, such as co-workers and customers, will be kept informed of expected coverage, deadlines and any updates,

Large organisations may have several schedules for different categories of news, for example General News, Sport, Finance, Features etc.

Increasingly, text and pictures are being augmented by dynamic content: video, audio, animated graphics, and the availability of this material needs to be signalled in the schedule to interested parties in a way that is amenable to software processing.

These business processes are addressed by **EventsML-G2** and **Editorial Planning – the Planning Item**.

6.3 Information Gathering

Most people recognise the model – beloved of Hollywood – of reporters, photographers, film/video and sound personnel rushing to the scene of a news event and generating content based on material they are able to obtain as the event unfolds.

In fact, news is gathered by an endless variety of means, such as press releases, reports from news agencies and freelance journalists, tip-offs from the public, statements on web sites, blogs etc. Generally, information gathered in this way is incomplete and needs to be augmented by additional material. Sometimes this material is gathered and prepared by contributors, working with the original creator.



This information gathering process ultimately results in journalists submitting event coverage: written copy, photographs, video footage and so on, to the Verification Stage of an editorial workflow.

6.4 Verification

The process of verifying the authenticity of news often starts before the content is generated, as part of the selection and assignment process. However, the detail of the content needs to be checked before the content can be released.

Responsible news organisations take steps to ensure that the facts of any news coverage are correct, and that they are presented in a fair, balanced and impartial way. It is also surprisingly easy to break the law by the inappropriate release of content. Lawyers or legally-trained staff routinely work with editors to ensure that content does not transgress the civil or criminal law, and that it is not gratuitously offensive to individuals or groups.

Clear and consistent writing, spelling and grammar are considered important and an organisation's rules will often be written down in a Style Guide which journalists are required to use when writing and editing.

Only when content meets all of the required standards will it be authorised for release. Completing these essential tasks under time pressure is one of the major operational challenges faced by news organisations.

6.5 Dissemination

Although seen conceptually as a physical "publication" process, the dissemination of information and news assets in digital form is pre-eminent today.

When news is received electronically, the recipient needs to be able to process the information quickly and reliably. When one considers that each day, a large news organisation may receive, from multiple sources, thousands of images, and hundreds of thousands of words of text, plus video, audio and graphics, the scale of the processing required becomes apparent.

The management of news requires organisations to know whether any given piece of content is useable, and in what context. Media organisations often receive content under *embargo*. This is information that has been released to professional journalists in advance so that they may complete any work needed to make it ready for dissemination to the public. Only when the *embargo time* has passed may the content be published. These informal protocols work because it is in the interests of all parties to co-operate. If they break an embargo, journalists know that their job may become more difficult because the provider will withhold information in future.

When content is transmitted electronically, it cannot be physically deleted by the provider. There must therefore be a means for providers to inform their customers that a piece of content must be deleted, ("cancelled" or "killed") and it is vital that any examples of the content are deleted from all systems, including archives, often for copyright or serious legal reasons.

The right to use a piece of content is an important aspect of news. Picture and video rights can be particularly complex. Although formal rights languages that are machine-readable, such as RightsML, (www.rightsml.org) are available, many organisations continue to indicate rights using a natural-language statement.

This management and administrative information must also be accompanied by descriptive information – metadata – that enables the receiver to direct the content to the appropriate workflow and users, retrieve related content, and if necessary re-purpose it for a variety of media channels.

Descriptive metadata will include some type of classification of the news so that its relevance to a sphere of interest(s) can be determined. Ad-hoc tags or keywords are useful, but their value is increased if they form part of a formal classification scheme, or taxonomy.

The use of taxonomies enables searches to yield consistent predictable results across a wide range of content and further enables accurate processing of content by software.



6.6 Archiving

A comprehensive digital archive of news, people and organisations plays an increasingly active role in the news process because of the features offered by electronic media such as the World Wide Web.

Today it is desirable to publish news which contains links to related news and information assets, allowing the consumer to view any aspect of a news story, including details of the people and organisations involved, and the concepts at issue.

The archiving process completes the news production cycle and accurate, comprehensive metadata is the key to unlocking the value of this information asset. The value of content is in direct proportion to the quality and quantity of its metadata; one can imagine that content with no metadata could be almost valueless.



This page intentionally blank

7 Quick Start: NewsML-G2 Basics

7.1 About the Quick Start Guides

This Quick Start Guide and its siblings in the next chapters are intended to give implementers enough information about NewsML-G2 to begin creating a useful working set of model NewsML-G2 documents for their organisation, or to begin working with G2 documents provided by another organisation. This Basics Guide covers the G2 features that are common to most types of content. Further Quick Guides give more specific information about conveying types of content: text, pictures, video, and news packages.

7.2 Introduction

The basic structure of a NewsML-G2 Item document is common to all applications. The available types of G2 Item are:

- ❖ News Item: for all kinds of news content.
- ❖ Package Item: for structured collections of news content.
- ❖ Concept Item: for expressing knowledge about entities, abstract concepts, and events.
- ❖ Knowledge Item: for collections of concepts, often grouped for a specific purpose such as Controlled Vocabularies.
- ❖ Planning Item: for exchanging information about news coverage and fulfilment.
- ❖ Catalog Item: for managing references to Controlled Vocabularies.

7.3 Item structure

The building blocks of a NewsML-G2 Item are shown in the diagram below:

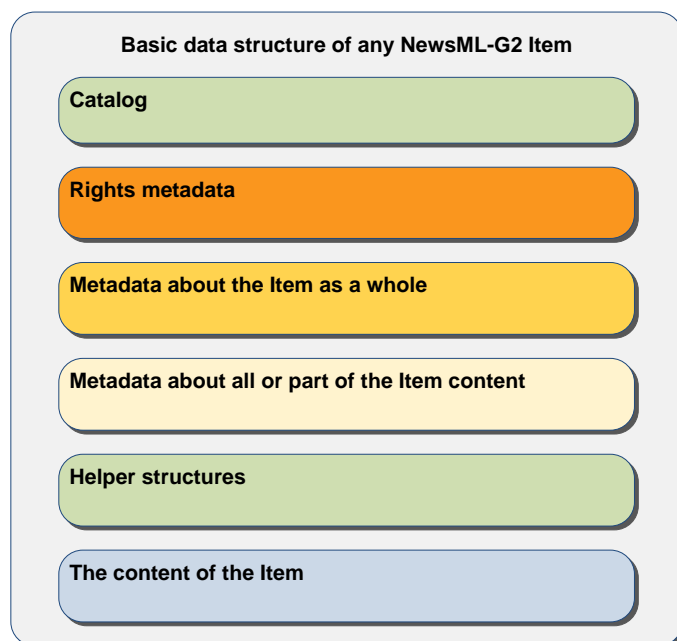


Figure 2: All G2 Items share this basic container structure

All have a root element that is specific to the type of Item that contains identification, version and some basic information to initiate the G2 processor.

The Catalog information is required to resolve QCodes, a fundamental feature of NewsML-G2 that enables partners to guarantee that codes used within an Item are globally unique.

The Rights Information block allow publishers to assert fine-grained information about copyright and usage terms as human-readable statements, or by inclusion of a machine-readable rights expression language such as RightsML.

The Item Metadata wrapper contains metadata about the Item as a whole, and this is followed by metadata about whole of the content `<contentMeta>`, and optionally by the `<partMeta>` wrapper, which enables publishers to express metadata about specific parts of the content..

Optional “helper” structures are available for specialised processing needs; their use is covered in **Chapter 25 Advanced Metadata Techniques**

Each type of G2 Item has a specific wrapper element for content, shown in the diagram below that also shows the basic top level elements common to all G2 Items:

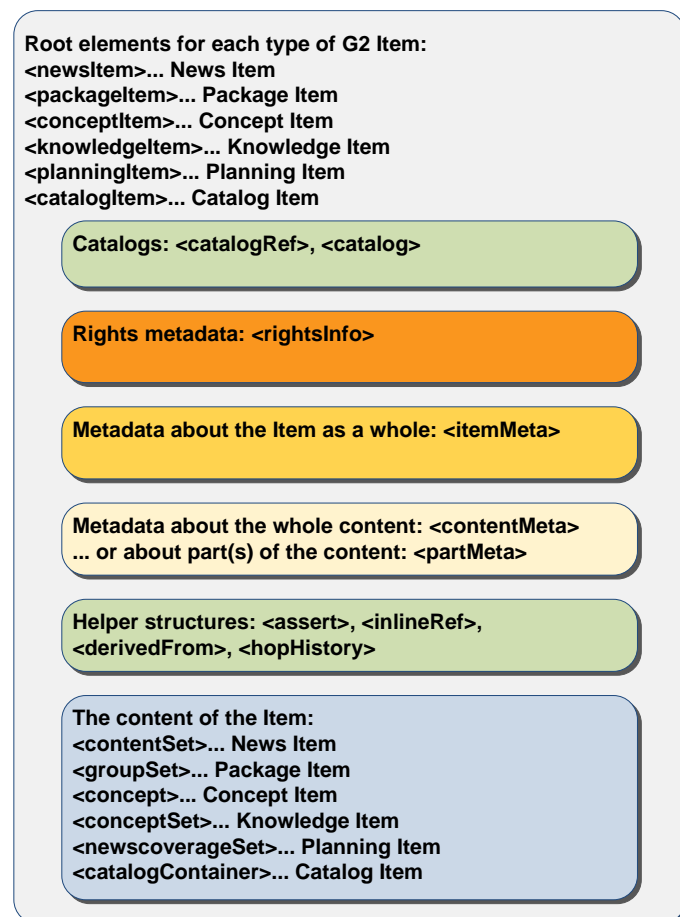


Figure 3: The basic XML elements associated with each part of a G2 Item



The example in this Quick Guide to G2 Basics uses a News Item with text content to illustrate the basic principles in action. Read this guide first, and proceed to further Quick Guides specific to Text, Pictures, Video, and Packages, as needed.

7.4 “Invisible” values

All NewsML-G2 documents contain some elements and attributes that have default or assumed values that allow the actual element or attribute to be omitted. These are detailed in a section entitled **Hidden Values of NewsML-G2** near the end of this Quick Start chapter.

7.5 Root element

See the diagram above for the root elements applicable to each G2 Item type.

7.5.1 Root element attributes

7.5.1.1 Item Identifier

All G2 Items must have a @guid, an identifier that should be globally unique for all time and independent of location. The IPTC has registered a URN namespace for the purpose of creating GUIDs for NewsML-G2 Items using a specification based on RFC3085. The syntax for a @guid using this scheme is:

| `guid="urn: newsml : [ProviderId] : [DateId] : [NewsItemId] "`

Use an internet domain name owned by your organisation as a ProviderId, for example:

| `<newsItem guid="urn: newsml : acmenews. com: 20141121: US- FINANCE- FED"`



Receivers of G2 Items must resist any temptation to “reverse engineer” the DateId part of a GUID to create a time-stamp. This may have unintended consequences and result in errors. Use the appropriate G2 timestamp property instead.

7.5.1.2 Version

A simple indicator of the version of the Item:

| `version="3"`

Version numbers need not be consecutive, but must START at 1, and a new version must be a higher number than the previous version. If version is missing, the value is assumed to be 1. See **Hidden Values of NewsML-G2**, below

7.5.1.3 Standard

A string denoting the IPTC Standard, in this case “NewsML-G2”.

| `standard="NewsML- G2"`

7.5.1.4 Standard Version

A string denoting the major and minor version of the Standard being used:

| `standardversion="2. 18"`

7.5.1.5 Conformance

There are two levels of Conformance to the NewsML-G2 standard. The “Core” conformance level (CCL) represents a minimal sub-set of G2 that can usefully get work done, and is the default if @conformance is omitted. In practice, many implementers use the properties of “Power” conformance (PCL). If an implementer stays within the CCL, the @conformance of the G2 Item can be assumed and omitted, but must be declared if using PCL:

| `conformance="power"`

7.5.1.6 Language

Setting the default language of XML elements with text content in NewsML-G2 is UK English:

| `xml : lang="en- GB"`

7.5.1.7 IPTC namespace

Sets the default namespace for elements:

| `xml ns="http: //iptc. org/std/nar/2006- 10- 01/"`

Putting this together, the required root element attributes are:

| `<newsItem xml ns="http: //iptc. org/std/nar/2006- 10- 01/"
guid="urn: newsml : acmenews. com: 20141121: US- FINANCE- FED"
version="3"`

```
standard="NewsML-G2"  
standardversion="2.18"  
conformance="power"  
xml:lang="en-GB">
```

7.5.2 Validating code

When developing a NewsML-G2 processing application, implementers will need to validate the generated G2 code against the appropriate schema. To validate code at PCL against the latest version of the NewsML-G2 specification (2.18) add the following code to the <newsItem> element:

```
xml ns="http://iptc.org/std/nar/2006-10-01/"  
xml ns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/  
http://www.iptc.org/std/NewsML-G2/2.18/specification/NewsML-G2_2.18-spec-All-  
Power.xsd"
```

To validate at CCL, substitute the name of the "Core" schema:

NewsML-G2_2.18-spec-All-Core.xsd



The IPTC does NOT permit validation of production documents against XML Schema files on the IPTC servers. The IPTC also recommends NOT validating production documents as this will create dependencies on external resources such as schema files and this could introduce points of failure in the processing of G2 Items.

7.5.3 Catalogs

Codes, short mnemonics used to express the value of properties such as Category, are a long-established feature of news exchange. QCodes are the NewsML-G2 mechanism that enables partners in news exchange to guarantee that codes are globally unique. Without going into details of the mechanism here, the News Item <catalog> enables a G2 processor to resolve QCodes, and guarantee that uniqueness, by mapping the code to a globally-unique URI. It is recommended that this URI locates a web resource.

One of the few mandatory NewsML-G2 elements, <itemClass>, uses QCodes issued by the IPTC to identify the business intention of the Item. For a News Item, the scheme is News Item Nature, with a recommended alias of "ninat". Values from the scheme include "ninat:text" and "ninat:picture". The catalog reference is:

```
<catalog>  
  <scheme alias="ninat" uri="http://cv.iptc.org/newscodes/nature/" />  
</catalog>
```

Other types of G2 Item use specific schemes for the <itemClass> property.

As the CVs used by a provider are usually quite consistent across the NewsML-G2 Items they publish, the IPTC recommends that the <catalog> references are aggregated into a stand-alone file which is made available as a web resource referenced by <catalogRef>. This is how the IPTC publishes its Catalogs:

```
<catalogRef  
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
```

The use of stand-alone web resources is preferable because all of the QCode mappings are shared across many G2 Items; a local <catalog> can only be used by the single Item.

It's likely that provider-specific catalogs will be needed to resolve QCodes used in the Item, for example:

```
<catalogRef  
  href="http://www.example.com/newsml-g2/catalog.eneews_2.xml" />
```

Adding the Catalog information to the example results in the following:

```
<newsItem xml ns="http://iptc.org/std/nar/2006-10-01/"  
  guid="urn:newsml:acmenews.com:20141121:US-FINANCE-FED"  
  version="3"  
  standard="NewsML-G2"  
  standardversion="2.18"  
  conformance="power"  
  xml:lang="en-GB">
```




```
<catalogRef  
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />  
<catalogRef  
  href="http://www.example.com/newsml-g2/catalog.eneews_2.xml" />
```

7.5.4 Rights information

The optional <rightsInfo> wrapper holds copyright information and usage terms, such as the following example:

```
<rightsInfo>  
  <copyrightHolder uri="http://www.example.com/about.html#copyright" >  
    <name>Example Enews LLP</name>  
  </copyrightHolder>  
  <copyrightNotice>  
    (c) Copyright 2014 Example Enews LLP, all rights reserved  
  </copyrightNotice>  
  <usageTerms>  
    Not for use outside the United States  
  </usageTerms>  
</rightsInfo>
```

7.6 Item Metadata <itemMeta>

7.6.1 Mandatory Properties

The mandatory <itemMeta> section has four mandatory elements, present in the following order:

7.6.1.1 Item Class

As previously mentioned, the <itemClass> property describes the type of content conveyed by the Item. It is mandatory to use one of the IPTC News Item Nature NewsCodes (recommended Scheme Alias: "ninat") for the Item Class of News Items and Package Items, expressed as a QCode:

```
<itemClass qcode="ninat:text" />
```

Possible values from this scheme include "ninat:picture", "ninat:video" and "ninat:audio".

7.6.1.2 Provider

Can be represented by a QCode, or a URI. If the value of this property is NOT taken from a controlled vocabulary, the @qcode or @uri will be omitted and the child <name> element used to give a human-readable value for the property. The IPTC recommends using a QCode with the Provider NewsCodes, a controlled vocabulary of providers registered with the IPTC with recommended Scheme Alias of "nprov";:

```
<provider qcode="nprov:REUTERS" />
```

7.6.1.3 Version Created

This contains the date, time and time zone (or UTC) that *this version* of the NewsML-G2 Item was created. The value must be expressed as XML Schema `dateTime`:
YYYY-MM-DDThh:mm:ss±hh:mm

```
<versionCreated>2014-11-21T16:25:32-05:00</versionCreated>
```

The -05:00 denotes U.S. Eastern Standard Time offset from UTC

7.6.1.4 Publication Status

Every G2 Item must have a publication status. The value defaults to "usable", which permits the <pubStatus> property to be omitted if desired, however it is recommended that a value is explicitly given:

```
<pubStatus qcode="stat:usable" />
```

Publication status is highly likely to be used by most news agencies, because the ability to explicitly signal the status of news is essential. The use of the IPTC Publishing Status NewsCodes is mandatory. Its recommended alias is "stat". Other values permitted by the scheme are:

- ❖ stat:canceled (note U.S. spelling) means that the content of the newsItem must not be used, ever.



- ❖ `stat:withheld` means the content must not be used until further notice.

When an Item is cancelled, it must never be used, but an Item that has been withheld may subsequently have its status updated to “usable”. For further information on processing rules, see section [26.2](#)

7.6.2 Optional Properties

The following optional properties are frequently used by G2 providers.

7.6.2.1 First Created

The `<firstCreated>` element indicates when first version of the Item (not the content) was first created.

```
<firstCreated>2010-10-18T13:12:21-05:00</firstCreated>
```

7.6.2.2 Embargoed

Business-to-business news organisations often use an *embargo* to release information in advance, on the strict understanding that it may not be released into the public domain until after the *embargo time* has expired, or until some other form of permission has been given.

Embargoed is NOT the same as the Publishing Status; embargoed content should have a publishing status of “usable”.

```
<embargoed>2014-10-23T12:00:00Z</embargoed>
```

It is not required to give any further information about the embargo conditions, but some providers may provide a natural language `<edNote>`, see below.

7.6.2.3 Service

The `<service>` element allows the provider to declare which of its services delivered this package, using a Controlled Vocabulary:

```
<service qcode="svc:uknews">  
  <name>UK News Service</name>  
</service>
```

7.6.2.4 Editorial Note

The `<edNote>` element contains a non-publishable note that is intended to be read by internal staff at the receiving organisation; in this example, some optional information about the release condition imposed by the `<embargoed>` element:

```
<edNote>  
  Note to editors: Updates previous version, adding more eye-witness reporting.  
</edNote>
```

7.6.2.5 Signal

Additional processing instructions can be given using `<signal>` and its `@qcode`. This example uses the IPTC Signal NewsCodes (recommended Scheme Alias “sig”) that advises the end-user that this Item updates a previous versions of the Item:

```
<signal qcode="sig:update" />
```

The other value in the Signal scheme is “correction”. There is further information on giving fine-grained information about signalling the reason and impact of updates in the section [26.6 Processing Updates and Corrections](#). For guidance on applying signals to different parts of an Item’s content, see [25.8](#)

7.6.2.6 Link

The `<link>` element has two basic purposes:

- ❖ To assert relationships to other Items, such as a previous version of an Item
- ❖ To create a navigable link from an Item to some supporting or additional resource.



This example provides a “see also” link to a resource on the Web that end-users can view to get further information about the event. @rel is used to denote the reason that the link is provided. In this example, the QCode uses the recommended IPTC Item Relation NewsCode with a recommended Scheme Alias of “irel” and the code value is “seeAlso”:

```
<link rel="irel:seeAlso"
href="http://www.example.com/video/20081222-PNN-1517-407624/index.html"/>
```

7.6.2.7 Completed Item Metadata

```
<itemMeta>
  <itemClass qcode="inat:text" />
  <provider qcode="nprov:REUTERS" />
  <versionCreated>2014-11-21T16:25:32-05:00</versionCreated>
  <firstCreated>2010-10-18T13:12:21-05:00</firstCreated>
  <embargoed>2014-10-23T12:00:00Z</embargoed>
  <pubStatus qcode="stat:usable" />
  <service qcode="svc:uknews">
    <name>UK News Service</name>
  </service>
  <edNote>
    Note to editors: STRICTLY EMBARGOED. Not for public release until 12noon
    on Wednesday, October 23.
  </edNote>
  <signal qcode="sig:update" />
  <link rel="irel:seeAlso"
    href="http://www.example.com/video/20081222-PNN-1517-407624/index.html"/>
</itemMeta>
```

7.7 Content Metadata <contentMeta>

Conceptually, there are two kinds of content metadata: Administrative and Descriptive.

7.7.1 Administrative Metadata

This is information about the content that cannot necessarily be deduced by examining it, for example: when it was created and/or modified. Administrative properties that are widely used by implementers of G2 for editorial content are:

- ❖ Timestamps (Created, Modified)
- ❖ Story location (Located)
- ❖ Creator
- ❖ Information Source

7.7.1.1 Timestamps

The <contentCreated> timestamp corresponds to a “Created on” field of the story. It is expressed in NewsML-G2 as Truncated DateTime data type, meaning that the date-time elements may optionally be stripped, starting from the right. If required, the <contentModified> property may also be used to contain the “Last Edit” timestamp. This must be later than the Created timestamp.

```
<contentCreated>2014-11-21T15:21:06-05:00</contentCreated>
<contentModified>2014-11-21T16:22:45-05:00</contentModified>
```

7.7.1.2 Located

The place that the content was created uses the <located> element, but note that this is not necessarily the place of the event or subject. For example, for a UK story written in the London office, <located> would be “London”; a picture of Mount Fuji taken from downtown Tokyo would have a <located> value of “Tokyo”.

The semantics of <located> are similar to the natural-language location carried in the dateline that often prefaces news (such as “BERLIN, October 24”) but can be conveyed more precisely, and in terms that may be more readily processed by software, using a @qcode or @uri:

```
<located type="cptype:city" qcode="geo:345678">
  <name>Berlin</name>
</located>
```

The optional @type uses a controlled vocabulary to indicate the nature of the location being expressed; in the example this indicates that <located> refers to a city.

7.7.1.3 Broader, Narrower

Both Located and Subject (below) contain child elements that express a specific relationship between entities or concepts. For example, the content originated in the city of Berlin and the <located> element shows that the city of Berlin has a “broader” relationship – that is a child-to-parent relationship – to Berlin the state, and to Germany the country:

```
<located type="cptype: city" qcode="geo: 345678">
  <name>Berlin</name>
  <broader type="cptype: statprov" qcode="prov: 2365">
    <name>Berlin</name>
  </broader>
  <broader type="cptype: country" qcode="iso3155-ia2: DE">
    <name>Germany</name>
  </broader>
</located>
```

7.7.1.4 Creator

The writer, photographer or other author of content is expressed using the <creator> element:

```
<creator uri="http://www.example.com/staff/mjameson" >
  <name>Meredith Jameson</name>
</creator>
```

7.7.1.5 Information Source

The <infoSource> element, together with its optional @role, enables finely-grained identification of the various parties who provided information used to create and develop an item of news. If absent, the default value of @role is the originator of the information used to create or enhance the content. For a detailed view on using <infoSource> see [27 Identifying Sources and Workflow Actors](#).

```
<infoSource uri="http://www.example.com/pressreleases/201411/newproducts.html" />
```

7.7.2 Descriptive Metadata

These properties set the context of news content in relation to other news by describing and classifying it. Information that has historically carried within the content itself, such as the headline and by-line (for text) or embedded metadata (pictures) may also be isolated as metadata. The practical benefit is that the end user no longer needs to scan or retrieve the actual content in order to process it. None of the Descriptive Metadata elements is mandatory, but the following feature frequently in G2 implementations.

7.7.2.1 Subject

The subject matter of content uses the <subject> element. When the value of the Subject is taken from a Controlled Vocabulary, this is identified using either a @qcode or @uri:

```
<subject type="cpnat: abstract" qcode="medtop: 20000523" />
```

or

```
<subject uri="http://subject.example.com/employment.html" />
```

For concepts not taken from a CV, the identifier is omitted and the name of the concept is given in the child <name> element, for example:

```
<subject>
  <name>Labour Market</name>
</subject>
```

The optional @type uses the IPTC “nature of the concept” NewsCodes (recommended scheme alias “cpnat”) to indicate the type of concept being expressed, for example, an abstract concept, that is a concept that does not represent a real-world entity, but something like an idea, or news category.

```
<subject type="cpnat: abstract" qcode="medtop: 04000000">
  <name xml:lang="en-GB">economy, business and finance</name>
```

`</subject>`

The above example uses a concept from the IPTC Media Topic NewsCodes. Also note the use of the W3C XML attribute `xml:lang` that expresses the language used for the element's value. It is also possible to add relationships to related concepts, as shown above in `<located>`. For example:

```
<subject type="cpnat: abstract" qcode="medtop: 04000000">
  <name xml:lang="en-GB">economy, business and finance</name>
</subject>
<subject type="cpnat: abstract" qcode="medtop: 20000523">
  <name xml:lang="en-GB">l labour market</name>
  <name xml:lang="de">Arbeitsmarkt</name>
  <broader qcode="medtop: 04000000" /> ← relationship property
</subject>
```



The IPTC highly recommends that providers use the Media Topic NewsCodes unless there is an over-riding requirement to use proprietary codes. This promotes inter-operability and standardisation in news exchange. It is also recommended that if a `<name>` is used in conjunction with a QCode, then the value of the `<name>` agrees with the value in the Scheme for the language specified in `xml:lang`. IPTC NewsCodes are provided in UK English (en-GB) and translations in French and German have been provided by IPTC members.

The “medtop” code prefix is the recommended Scheme Alias for the Codes, which is resolved via the IPTC Catalog (see `<catalogRef>`, above) to the Scheme URI <http://cv.iptc.org/newscodes/mediatopic/>. This is guaranteed to be globally unique because it is part of the Internet Domain controlled by the IPTC. Appending the code “04000000” to the Scheme URI forms a Concept URI that cannot be confused with a concept with the same code from another source.

When using `@type` to indicate the nature of the concept, the possible values from the IPTC Concept Nature NewsCodes are:

- ❖ Abstract: a concept that does not represent a real-world entity
- ❖ Event
- ❖ geoArea: a geo-political area
- ❖ Object: A real-world object, such as a painting; an aircraft
- ❖ Organisation
- ❖ Person
- ❖ Point of Interest

7.7.2.2 Genre

The `<genre>` element indicates the style of the content, in this example “interview” as a property that is distinct from `<subject>` that is used to indicate the subject matter of content. In the example, the IPTC Genre NewsCode for “Interview” is used:

```
<genre qcode="genre: interview">
  <name xml:lang="en-GB">Interview</name>
</genre>
```

7.7.2.3 Slugline



Some news services implemented in G2 retain the `<slugline>` property as a human-readable index for legacy reasons; therefore receivers may sometimes see this property. However, it has never been a completely reliable identifier, and it is recommended that more purposeful identifiers that are also machine-readable are implemented in its place.

```
<slugline>US- Finance- Fed</slugline>
```

7.7.2.4 Headline

Even if the Headline is carried inline in text content, it is useful to place it in metadata so that it can more easily be identified and extracted by the end-user:

```
<headline>Fed to halt QE to avert “bubble”</headline>
```



7.7.2.5 Completed Content Metadata

```
<contentMeta>
  <contentCreated>2014-11-21T15:21:06-05:00</contentCreated>
  <contentModified>2014-11-21T16:22:45-05:00</contentModified>
  <located type="cptype:city" qcode="geo:345678">
    <name>Berlin</name>
    <broadier type="cptype:statprov" qcode="prov:2365">
      <name>Berlin</name>
    </broadier>
    <broadier type="cptype:country" qcode="iso3155-ia2:DE">
      <name>Germany</name>
    </broadier>
  </located>
  <creator uri="http://www.example.com/staff/mjameson" >
    <name>Meredith Jameson</name>
  </creator>
  <infoSource uri="http://www.example.com" />
  <subject type="cpnat:abstract" qcode="medtop:04000000">
    <name xml:lang="en-GB">economy, business and finance</name>
  </subject>
  <subject type="cpnat:abstract" qcode="medtop:20000523">
    <name xml:lang="en-GB">labour market</name>
    <name xml:lang="de">Arbeitsmarkt</name>
    <broadier qcode="medtop:04000000" />
  </subject>
  <genre qcode="genre:interview">
    <name xml:lang="en-GB">Interview</name>
  </genre>
  <slugline>US-Finance-Fed</slugline>
  <headline>Fed to halt QE to avert "bubble"</headline>
</contentMeta>
```

7.8 Content

The content of a NewsML-G2 document varies according to its Item type: the example below shows a News Item with a <contentSet> wrapper containing a trivial text payload. Following this code example are skeletal examples showing the other options for conveying content in G2 News Items and Package Items:

LISTING 1 A NewsML-G2 News Item

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:acmenews.com:20141121:US-FINANCE-FED"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-GB">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef href="http://www.example.com/newsml-g2/catalog.eneews_2.xml" />
  <rightsInfo>
    <copyrightHolder uri="http://www.example.com/about.html#copyright" >
      <name>Example Enews LLP</name>
    </copyrightHolder>
    <copyrightNotice>
      (c) Copyright 2014 Example Enews LLP, all rights reserved
    </copyrightNotice>
    <usageTerms>
      Not for use outside the United States
    </usageTerms>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ninat:text" />
    <provider qcode="nprov:REUTERS" />
    <versionCreated>2014-11-21T16:25:32-05:00</versionCreated>
    <firstCreated>2010-10-18T13:12:21-05:00</firstCreated>
    <embargoed>2014-10-23T12:00:00Z</embargoed>
    <pubStatus qcode="stat:usable" />
    <service qcode="svc:uknews">
      <name>UK News Service</name>
    </service>
    <edNote>
      Note to editors: STRICTLY EMBARGOED. Not for public release until 12noon
      on Wednesday, October 23.
    </edNote>
```



```
<signal qcode="sig:update" />
<link rel="irel:seeAlso"
href="http://www.example.com/video/20081222-PNN-1517-407624/index.html"/>
</itemMeta>
<contentMeta>
  <contentCreated>2014-11-21T15:21:06-05:00</contentCreated>
  <contentModified>2014-11-21T16:22:45-05:00</contentModified>
  <located type="cptype:city" qcode="geo:345678">
    <name>Berlin</name>
    <broadier type="cptype:statprov" qcode="prov:2365">
      <name>Berlin</name>
    </broadier>
    <broadier type="cptype:country" qcode="iso3155-ia2:DE">
      <name>Germany</name>
    </broadier>
  </located>
  <creator uri="http://www.example.com/staff/mjameson" >
    <name>Meredith Jameson</name>
  </creator>
  <infoSource uri="http://www.example.com" />
  <subject type="cpnat:abstract" qcode="medtop:04000000">
    <name xml:lang="en-GB">economy, business and finance</name>
  </subject>
  <subject type="cpnat:abstract" qcode="medtop:20000523">
    <name xml:lang="en-GB">labour market</name>
    <name xml:lang="de">Arbeitsmarkt</name>
    <broadier qcode="medtop:04000000" />
  </subject>
  <genre qcode="genre:interview">
    <name xml:lang="en-GB">Interview</name>
  </genre>
  <slugline>US-Finance-Fed</slugline>
  <headline>Fed to halt QE to avert "bubble"</headline>
</contentMeta>
<contentSet>
  <inlineXML contenttype="application/nitf+xml"> <!-- A VALID MIME TYPE -->
    <!-- Inline XML must contain well-formed XML such as NITF or XHTML -->
  </inlineXML>
</contentSet>
</newsItem>
```

7.8.1 News Content options for News Items and Package Items

The News Item <contentSet> contains a single logical piece of content, but allows alternative renditions of the SAME content to be carried in a single G2 Item:

```
<contentSet>
  <inlineXML contenttype="application/nitf+xml"> <!-- A VALID MIME TYPE -->
    <!-- Inline XML must contain well-formed XML such as NITF or XHTML -->
  </inlineXML>
</contentSet>
```

or

```
<contentSet>
  <inlineData>
    <!-- Inline Data contains plain text -->
  </inlineData>
</contentSet>
```

or

```
<contentSet>
  <!-- Remote Content contains reference to binary asset such as -->
  <!-- PDF or image file -->
  <remoteContent rendition="rnd:one" />
  <remoteContent rendition="rnd:two" />
  <remoteContent rendition="rnd:three" />
</contentSet>
```

Package Item content is wrapped by the <groupSet> element, which can contain one or more <group> children. Each Group contains references to Items that make up the News Package, using the <itemRef> element. A Group can also reference other Groups via the <groupRef> element.

```
<groupSet ... >
  <group ... >
```



```

    <itemRef ... >
    <!-- Properties extracted from the packaged Item -->
    </itemRef>
    <groupRef />
  </group>
  <group ... >
    ...
  </group>
</groupSet>

```

7.9 Summary and Next Steps

This section has covered the basic structure that is common to all G2 Items, and also outlined properties that are commonly used for news content. Further Quick Guides show how to build upon this foundation:

Quick Start – Text takes an example news story and shows how the information on an editor’s screen would be implemented in NewsML-G2.

Quick Start – Pictures takes an example image and its embedded metadata and converts this to a G2 properties with several image renditions carried in a single G2 Item. The guide also shows how to express the various technical characteristics of images using G2 properties.

Quick Start – Video is split into two sections: the first covers a simple case of a standalone video file with various technical renditions expressed in G2; the second uses a more comprehensive structure that separates the metadata for multiple segments of a video, using the `<partMeta>` wrapper.

Quick Start – Packages shows how G2 Items and other kinds of content can be assembled into Packages of managed objects with an explicit structure.

7.10 Hidden Values of NewsML-G2

There are some default values set by the specification which allow an element or attribute to be omitted and the default assumed. The list below shows NewsML-G2 elements and attributes which optionally appear in an Item but for which a usable value or status exists.

7.10.1 All G2 items

- ❖ @version of the root element = “1”
- ❖ @conformance of the root element = “core”
- ❖ <embargoed> = no embargo
- ❖ <pubStatus> = “usable”
- ❖ <catalog> and <catalogRef> = one is required; many of either, or a mix of both may be used
- ❖ @scope of <hash> = “content”. Hash value is a message digest included in the Item for security purposes. A hash scope of “content” indicates that the hash value was derived by hashing the content only.
- ❖ @how (attribute of many elements) = “person”. The attribute indicates how the value was extracted from the content: by a person
- ❖ @custom (attribute of many elements) = “false”. The attribute indicates that the property was added specifically for a customer or group of customers
- ❖ @dir (many elements) = “ltr”. The directionality of the script of the language of the property is left to right.

7.10.2 News Items only

- ❖ @timeunit (when @duration is used) = “seconds”
- ❖ @dimensionunit (when @width/@height is used) = (for example) pixels for a still picture. See the Quick Start Guide – Pictures for details.
- ❖ @encoding of <inlineData> = “base64”. This is the default binary encoding for Inline Data.

7.10.3 Package Item only

- ❖ @mode of the <group> = “bag” – an unordered collection of complementary components

8 Quick Start: Text

8.1 Introduction

One of the most fundamental needs of a news organisation is to handle text. This chapter covers the basics of a simple NewsML-G2 News Item containing text content.



Read the Quick Guide to G2 Basics
before this Quick Start

8.2 Example

Below is an example story and supporting information as might be displayed on the journalist's editing screen at a fictional news provider, Acme News and Media (ANM):

Acme News and Media – Content Editing System		
Slugline	US-Finance-Fed	
Created on	2014-11-21 15:21:06	
Source	ANM	
Author	mjameson	
Latest edit	2014-11-21 16:22:45	
Latest editor	moiras	
Categories	economy, finance, business, central bank, monetary policy	
Headline	Fed to halt QE to avert "bubble"	
Byline	By Meredith Jameson	
Location / Date	Washington	21/11/2014
Body Text	Et, sent luptat luptat, commy nim zzriureet vendreetue modo dolenis ex eisis nosto et lan ullandit lum doloreet vulla feugiam coreet, cons eleniam il ute facin veril et aliquis ad minis et lor sum del iriure dit la feugiamcommy nostrud min ullapat velisl duisismodip ero dipit nit utpatum sandrer cipisim nit lortis augiat nulla faccum at am, quam velenis nulpur la auguerostrud magna commolore eliquatie exerate facilis modiamconsed dion henisse quipit at. Ut la feu facilla feu faccumsan ecte modoloreet ad ex el utat.	

This screen contains nearly all of the information needed to create a valid NewsML-G2 document.

8.3 Document structure

The building blocks of the text document are the <newsItem> root element, with additional wrapping elements for metadata about the News Item (itemMeta), metadata about the content (contentMeta) and the content itself (contentSet). The top level (root) element <newsItem> attributes are:

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:acmenews.com:20141121:US-FINANCE-FED"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-US">
```

This is followed by references to the Catalogs used to resolve QCodes in the Item, and Rights information:

```
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<catalogRef
  href="http://catalog.acmenews.com/news/ANM_G2_CODES_2.xml" />
<rightsInfo>
  <copyrightHolder uri="http://www.acmenews.com/about.html#copyright" >
    <name>Acme News and Media LLC</name>
  </copyrightHolder>
  <copyrightNotice>(c) 2014 Copyright Acme News and Media LLC</copyrightNotice>
</rightsInfo>
```

8.3.1 Item Metadata <itemMeta>

Note the three mandatory child elements of the mandatory <itemMeta>:

- ❖ Item Class
- ❖ Provider
- ❖ Version Created

In addition, a publication status is also mandatory, but the <pubStatus> element that expresses it may be omitted if the publication status is "usable", the default. It is recommended that the publication status is explicitly carried as in this example. As Acme News & Media is fictional, the Provider property does not use one of the IPTV Provider NewsCodes, and so it is expressed using a URI:

```
<itemMeta>
  <itemClass qcode="inat:text" />
  <provider uri="http://www.acmenews.com/about.html" />
  <versionCreated>2014-11-21T16:25:32-05:00</versionCreated>
  <pubStatus qcode="stat:usable" />
  <embargoed>2014-11-26T12:00:00-05:00</embargoed>
</itemMeta>
```

8.3.2 Content Metadata <contentMeta>

8.3.2.1 Administrative Metadata

The administrative properties of the example text story are:

```
<contentCreated>2014-11-21T15:21:06-05:00</contentCreated>
<contentModified>2014-11-21T16:22:45-05:00</contentModified>
```

The place that the content was created uses the <located> element:

```
<located qcode="geoloc:NYC">
  <name>New York, NY</name>
</located>
```

(Note that this is where the story was written, not the place where the subject of the story took place. That would be expressed using <subject>, part of Descriptive Metadata.)

The author of the article is expressed using the <creator> element:

```
<creator uri="http://www.acmenews.com/staff/mjameson">
  <name>Meredith Jameson</name>
</creator>
```

The Information Source for the article is also given. When used without a @role, <infoSource> is used to denote the person or party that provided the original information on which the content is based. This is the relationship to be expressed here:

```
<infoSource qcode="is:AP">
  <name>Associated Press</name>
</infoSource>
```

The default language for the content is given as U.S. English:

```
<language tag="en-US" />
```

8.3.2.2 Descriptive Metadata

In the example, the Subject properties use QCodes from the Controlled Vocabulary of Media Topics NewsCodes that are owned and maintained by the IPTC and expressed as QCodes. Thus:

```
<subject qcode="medtop:04000000">
  <name>economy, business and finance</name>
</subject>
<subject qcode="medtop:20000350">
  <name>central bank</name>
</subject>
<subject qcode="medtop:20000379">
  <name>money and monetary policy</name>
</subject>
```

The <slugline> property contains the value of the “Slugline” field of the story:

```
<slugline>US- Finance- Fed</slugline>
```

In a similar fashion, the <headline> property will contain the value of the “Headline” field:

```
<headline>Fed to halt QE to avert “bubble”</headline>
```

8.3.2.3 Complete Content Metadata

```
<contentMeta>
  <contentCreated>2014- 11- 21T15: 21: 06- 05: 00</contentCreated>
  <contentModified>2014- 11- 21T16: 22: 45- 05: 00</contentModified>
  <located qcode=“geoloc: NYC”>
    <name>New York, NY</name>
  </located>
  <creator uri=“http://www.acmenews.com/staff/mjameson”>
    <name>Meredith Jameson</name>
  </creator>
  <infoSource qcode=“is: AP”>
    <name>Associated Press</name>
  </infoSource>
  <language tag=“en- US” />
  <subject qcode=“medtop: 04000000”>
    <name>economy, business and finance</name>
  </subject>
  <subject qcode=“medtop: 20000350”>
    <name>central bank</name>
  </subject>
  <subject qcode=“medtop: 20000379”>
    <name>money and monetary policy</name>
  </subject>
  <slugline>US- Finance- Fed</slugline>
  <headline>Fed to halt QE to avert “bubble”</headline>
</contentMeta>
```

8.4 Text content choices

8.4.1 Inline XML

The content of the NewsML-G2 document is enclosed by the <contentSet> wrapper. In the example, the IPTC news mark-up language NITF (News Industry Text Format) is used to format the text content. As an XML standard, it is contained in an <inlineXML> child element of <contentSet>, and uses @contenttype to denote the XML-based standard, using the IANA MIME type.

XHTML is also a popular text mark-up choice among G2 providers. As alternatives, the contents of <inlineXML> may be any XML language that can express generic or specialised news information, including SportsML-G2 and EventsML-G2. Other languages such as XBRL (Extended Business Reporting Language) may also be used. The content inside <inlineXML> must be valid XML, in other words, it could stand alone as a valid XML document in its own namespace.

```
<contentSet>
  <inlineXML contenttype=“application/nitf+xml”>
    <nitf xmlns=“http://iptc.org/std/NITF/2006- 10- 18/”>
      <!-- STORY CONTENT HERE -->
    </nitf>
  </inlineXML>
</contentSet>
```

8.4.2 Inline data

The <inlineData> element can contain plain text, and in this case MUST be identified by the IANA MIME type of “text/plain” thus:

```
<contentSet>
  <inlineData contenttype=“text/plain”>
    Et, sent luptat luptat ...
  </inlineData>
</contentSet>
```



8.5 Putting it together

LISTING 2 NewsML-G2 Text Document

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:acmenews.com:20141121:US-FINANCE-FED"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef
    href="http://catalog.acmenews.com/news/ANM_G2_CODES_2.xml" />
  <rightsInfo>
    <copyrightHolder uri="http://www.acmenews.com/about.html#copyright">
      <name>Acme News and Media LLC</name>
    </copyrightHolder>
    <copyrightNotice>(c) 2014 Copyright Acme News and Media LLC</copyrightNotice>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="inat:text" />
    <provider uri="http://www.acmenews.com/about/" />
    <versionCreated>2014-11-21T16:25:32-05:00</versionCreated>
    <embargoed>2014-11-26T12:00:00-05:00</embargoed>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <contentCreated>2014-11-21T15:21:06-05:00</contentCreated>
    <contentModified>2014-11-21T16:22:45-05:00</contentModified>
    <located qcode="geoloc:NYC">
      <name>New York, NY</name>
    </located>
    <creator uri="http://www.acmenews.com/staff/mjameson">
      <name>Meredith Jameson</name>
    </creator>
    <infoSource qcode="is:AP">
      <name>Associated Press</name>
    </infoSource>
    <language tag="en-US" />
    <subject qcode="medtop:04000000">
      <name>economy, business and finance</name>
    </subject>
    <subject qcode="medtop:20000350">
      <name>central bank</name>
    </subject>
    <subject qcode="medtop:20000379">
      <name>money and monetary policy</name>
    </subject>
    <slugline>US-Finance-Fed</slugline>
    <headline>Fed to halt QE to avert "bubble"</headline>
  </contentMeta>
  <contentSet>
    <inlineXML contentType="application/nitf+xml">
      <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/">
        <body>
          <body.head>
            <headline>
              <hl1>Fed to halt QE to avert "bubble"</hl1>
            </headline>
            <byline>By Meredith Jameson, <byttl>Staff Reporter</byttl></byline>
          </body.head>
          <body.content>
            <p>(New York, NY - November 21) Et, sent luptat luptat, commy Nim
              zzriureet vendreetue modo dolenis ex euisis nosto et lan ullandit
              lum doloreet vulla.
            </p>
            <p>Ugiating ea feugait utat, venim velent nim quis nulluptat num
              Volorem inci enim dolobor eetuer sendre ercin utpatio dolorpercing.
            </p>
          </body.content>
        </body>
      </nitf>
    </inlineXML>
  </contentSet>
</newsItem>
```

9 Quick Start: Pictures and Graphics

9.1 Introduction

Image content, including pictures and graphics, can be conveyed in a standard NewsML-G2 document. Picture providers and consumers need a rich vocabulary for descriptive and technical metadata, and for administrative metadata such as rights and usage terms. There is also a long-established use of embedded metadata, such as the IPTC/IIM Fields in JPEG and TIFF files. This Quick Start guide addresses some aspects of embedded metadata, but for a full description of the mapping embedded metadata and IIM fields to G2, this is described in detail in [22 Mapping Embedded Photo Metadata to NewsML-G2](#).

The example in this Quick Start guide is a simple but complete document that shows how to implement in NewsML-G2 the frequently-used needs of a professional picture workflow:

- ❖ Right and usage instructions
- ❖ Descriptive and administrative properties such as Location and Categorisation
- ❖ Separate technical renditions of a picture



Read the Quick Guide to G2 Basics
before this document.

The picture and the metadata used in the example are courtesy of Getty Images. *Note that the sample code is NOT intended to be a guide to receiving NewsML-G2 from Getty Images.*

9.2 About the example



A library picture is provided to customers in three sizes: a large image intended for high resolution and/or large size display, a medium-sized image intended for web use, and a small image for use as a thumbnail or icon. These are three *alternative* renditions of the same picture and can be contained in a single NewsML-G2 document.

9.3 Document structure

The building blocks of the NewsML-G2 document are the <newsItem> root element, with additional wrapping elements for metadata about the News Item (itemMeta), metadata about the content (contentMeta) and the content itself (contentSet).

The root <newsItem> attributes are:

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"  
  guid="tag:gettyimages.com.2010:GYI0062134533"  
  version="3"  
  standard="NewsML-G2"  
  standardversion="2.18"  
  conformance="power"  
  xml:lang="en-US">
```

Note that this example uses a Tag URI (see [TAG URI home page](#) for details)

This is followed by references to the Catalogs used to resolve QCodes in the Item, and Rights information:

```
<catalogRef  
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />  
<catalogRef href="http://cv.gettyimages.com/nml2catalog4customers-1.xml" />  
<rightsInfo>  
  <copyrightHolder uri="http://www.gettyimages.com">  
    <name>Getty Images North America</name>  
  </copyrightHolder>  
  <copyrightNotice  
    href="http://www.gettyimages.com/Corporate/LicenseInfo.aspx">  
    (c) 2010 Copyright Getty Images. --  
    http://www.gettyimages.com/Corporate/LicenseInfo.aspx  
  </copyrightNotice>  
  <usageTerms>Contact your local office for all commercial or  
    promotional uses. Full editorial rights UK, US, Ireland, Canada (not  
    Quebec). Restricted editorial rights for daily newspapers elsewhere,  
    please call.</usageTerms>
```



| </rightsInfo>

9.3.1 Source

Note that the IIM "Source" field maps to the G2 <copyrightHolder> element of the <rightsInfo> block.

9.4 Item Metadata <itemMeta>

```
<itemMeta>
  <itemClass qcode="inat:picture">
    <provider qcode="nprov:GYI">
      <name>Getty Images Inc. </name>
    </provider>
    <versionCreated>2014-11-12T06:42:04Z </versionCreated>
    <firstCreated>2010-10-20T20:58:00Z</firstCreated>
  </itemMeta>
```

The <itemClass> property uses a QCode from the IPTC News Item Nature NewsCodes to denote that the Item conveys a picture.

The Z suffix denotes UTC. Note the <firstCreated> property refers to the creation of the Item, NOT the content.

9.5 Embedded metadata

For many years IPTC metadata fields have been embedded in JPEG or TIFF image files. From 1995 on the IPTC Information Interchange Model (IIM) defined the semantics of the fields and the technical format for saving them in image files. In 2003 Adobe introduced a new format for saving metadata, namely XMP (Extended Metadata Platform), and many IPTC IIM fields were specified as the "IPTC Core" metadata schema. This defined identical semantics but opened the formats for saving to IIM and XMP in parallel. Later the "IPTC Extension" metadata schema was added; the defined fields are stored by XMP only. Thus, many people work with IPTC photo metadata, regardless how they are saved in the files; this is handled by the software they use.

The transfer of IPTC Photo Metadata fields to NewsML-G2 properties has a focus on the equivalence of the semantics of fields. The retrieval of the embedded values from the files is a secondary issue and documents like the Guidelines for Handling Image Metadata, produced by the Metadata Working Group (<http://www.metadataworkinggroup.org/specs/>) help in this area.

This Quick Start guide will provide the basics of this mapping, for more details see **22 Mapping Embedded Photo Metadata to NewsML-G2**, and see the IPTC web site for more about the IPTC Core and Extension Schemas for XMP (visit www.iptc.org and navigate to Photo Metadata).

The screen shot on the following page shows the panel for the IPTC Core fields as displayed by Adobe's Photoshop CS File Info screen; note the IPTC Extension tab that displays the additional IPTC Extension metadata.

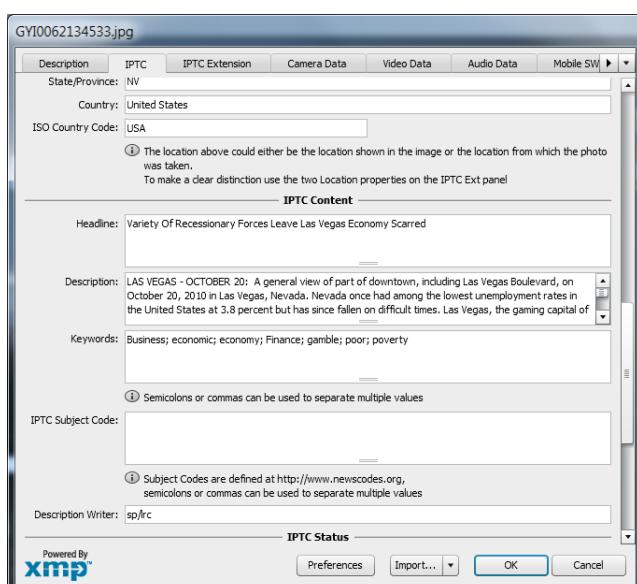


Figure 4: IPTC Core Metadata fields in the File Info panel of Adobe Photoshop

There are advantages, in a professional workflow, to carrying metadata independently of the binary asset:

- ❖ There is no need to retrieve and open the file to read essential information about the picture
- ❖ An editor may not have access to the original picture to modify its metadata
- ❖ A library picture used to illustrate a news event may have inappropriate embedded metadata.

A situation may arise where the metadata expressed in the NewsML-G2 Item and the embedded metadata in the photo are different. Some providers choose to strip all embedded metadata from objects, to avoid potential confusion. If not, a provider should specify any processing rules in its terms of use.

The IPTC recommends that *descriptive* metadata properties that exist in the NewsML-G2 Item (in Content Metadata) ALWAYS take precedence over the equivalent embedded metadata (if it exists). These properties include genre, subject, headline, description and creditline.

9.6 Content Metadata <contentMeta>

This example shows how embedded metadata from the example picture are translated into NewsML-G2, and includes the equivalent IPTC Core metadata schema property highlighted thus:

IPTC Core Schema equivalent:

9.6.1 Administrative metadata

9.6.1.1 Timestamp

The <contentCreated> element is used to give the creation date of the picture:

```
<contentCreated>2010- 10- 20T19: 45: 58- 04: 00</contentCreated>
```

Note that this value refers to the creation of the original content; for a scanned picture this is always the date (and optionally the time) of the original photograph. The property type is Truncated Date Time, so that when the precise date-time is unknown, for example for an historic photograph, the value can be truncated (from the right) to a simple date or just a year.

IPTC Core Schema equivalent: Date Created

9.6.1.2 Creator

The example uses a <creator> element without an identifier, but includes an optional @role that contains a QCode qualifying the creator as a photographer:

```
<creator role="crol: photographer">
  <name>Spencer Platt</name>
  <related rel="crol: isa" qcode="gyi bt: staff" />
</creator>
```



The <related> child element of <creator> further qualifies the photographer as a member of staff (as distinct from, say, a freelance photographer)

IPTC Core Schema equivalent: Creator

9.6.1.3 Contributor

A <contributor> identifies people or organisations who did not originate the content, but have added value to it. In this case, the @role hints that the contributor added descriptive metadata:

```
<contributor role="ctrol:descrWriter">
  <name>sp/1rc</name>
</contributor>
```

IPTC Core Schema equivalent: Description Writer

9.6.1.4 Creditline

The <creditline> is a natural-language string that must be used by the receiver to indicate the credit(s) for the content, as directed in the business terms agreed with the provider or copyright holder:

```
<creditline>Getty Images</creditline>
```

IPTC Core Schema equivalent: Credit Line

9.6.2 Descriptive metadata

9.6.2.1 Subject

As described in the Quick Start Guide to G2 Basics, the subject matter of content is expressed using the <subject> element. The optional @type uses the IPTC Concept Nature NewsCodes (recommended scheme alias "cpnat") to indicate the type of concept being expressed. The following example uses a value of "cpnat:event" to indicate that the concept is an Event, and the QCode identifies the Event in the scheme with an alias "gyimeid":

```
<subject type="cpnat: event" qcode="gyimeid: 104530187" />
```

The provider can use this Event ID to "tag" each of the pictures that relate to this topic, enabling receivers to group them via the Event ID.

The picture of Las Vegas Boulevard illustrates a story about unemployment. This example uses codes and associated <name> child elements from the IPTC Media Topic NewsCodes:

```
<subject type="cpnat: abstract" qcode="medtop: 20000523">
  <name xml:lang="en-GB">l abour market</name>
  <name xml:lang="de">Arbei tsmarkt</name>
</subject>
<subject type="cpnat: abstract" qcode="medtop: 20000533">
  <name xml:lang="en-GB">unempl oyment</name>
  <name xml:lang="de">Arbei tsi gkei t</name>
</subject>
```

9.6.2.2 City, State/Province, Country

The <located> element in the <contentMeta> block describes the place where the picture was created. This may be the same location as the event portrayed in the picture, but this cannot be assumed. The location of the event is logically part of the subject matter – the City, State/Province, Country fields in the IPTC Photo Metadata are defined as "the location shown" – so should use the <subject> element. To summarise:

- ❖ Use <located> to describe where the camera was located when taking the picture.
- ❖ Use <subject> to describe the location shown in the picture. It is recommended that @type is used to indicate the property identifies a geographical area.

The location shown in the example picture is Las Vegas Boulevard. Child elements of <subject> may be used to add further details, including:

- ❖ <name> gives the place name in plain text, and



- ❖ `<broadener>`² expresses the concept of Las Vegas Boulevard as part of the broader entity of Las Vegas which in turn is part of broader entities of Nevada state and of the United States.

It is recommended that the nature of the concept is indicated by `@type` using a value from the IPTC Concept Nature NewsCodes, in this case that the concept identifies a geographical area:

The completed `<subject>` structure for the geographical information is:

```
<subject type="cpnat: geoArea">
  <name>Las Vegas Boulevard</name>
</subject>
<subject type="cpnat: geoArea" qcode="gycon: 89109">
  <name>Las Vegas</name>
  <broadener qcode="iso3166- 2: US- NV">
    <name>Nevada</name>
  </broadener>
  <broadener qcode="iso3166- 1a3: USA">
    <name>United States</name>
  </broadener>
</subject>
```

9.6.2.3 Keywords

QCodes and relationship properties are powerful tools, but keywords are still widely used by picture archives. The G2 `<keyword>` property is mapped from the “Keywords” field in XMP. The semantics of “keyword” can vary from provider to provider, but should not present problems in the news industry, which is familiar enough with their use:

```
<keyword>business</keyword>
<keyword>economic</keyword>
<keyword>economy</keyword>
<keyword>finance</keyword>
<keyword>poor</keyword>
<keyword>poverty</keyword>
<keyword>gamble</keyword>
```

IPTC Core Schema equivalent: Keywords

9.6.2.4 Headline, Description

These two IPTC/IIM fields map directly to elements of the same name in NewsML-G2. Both `<headline>` and `<description>` also have an optional `@role`. The IPTC maintains a set of NewsCodes for Description Role (recommended scheme alias “drol”). In this case, as the description is of a photograph, the role will be “caption”. Description is a *Block* type element, meaning it may contain line breaks.

Both elements have optional attributes which may be used to support international use: `@xml:lang`, `@dir` (text direction):

```
<headline>Variety Of Recessionary Forces Leave Las Vegas
  Economy Scarred</headline>
<description role="drol: caption">A general view of part of downtown,
  including Las Vegas Boulevard, on October 20, 2010 in Las Vegas,
  Nevada. Nevada once had among the lowest unemployment rates in the
  United States at 3.8 percent but has since fallen on difficult times.
  Las Vegas, the gaming capital of America, has been especially hard
  hit with unemployment currently at 14.7 percent and the highest
  foreclosure rate in the nation. Among the sparkling hotels and
  casinos downtown are dozens of dormant construction projects and
  hotels offering rock bottom rates. As the rest of the country slowly
  begins to see some economic progress, Las Vegas is still in the midst
  of the economic downturn. (Photo by Spencer Platt/Getty Images)
</description>
```

IPTC Core Schema equivalent: Headline

9.6.3 Completed `<contentMeta>`

```
<contentMeta>
  <contentCreated>2010- 10- 20T19: 45: 58- 04: 00</contentCreated>
```

² `<broadener>` is only available at Power Conformance Level, which is why we set `@conformance` to “power” in `<newsItem>`

```

<creator role="crol:photographer">
  <name>Spencer Platt</name>
  <related rel="crel:isa" qcode="gyibt:staff" />
</creator>
<contributor role="ctrol:descWriter">
  <name>sp/lrc</name>
</contributor>
<creditline>Getty Images</creditline>
<subject type="cpnat:event" qcode="gyimeid:104530187" />
<subject type="cpnat:abstract" qcode="medtop:20000523">
  <name xml:lang="en-GB">labour market</name>
  <name xml:lang="de">Arbeitsmarkt</name>
</subject>
<subject type="cpnat:abstract" qcode="medtop:20000533">
  <name xml:lang="en-GB">unemployment</name>
  <name xml:lang="de">Arbeitslosigkeit</name>
</subject>
<subject type="cpnat:geoArea">
  <name>Las Vegas Boulevard</name>
</subject>
<subject type="cpnat:geoArea" qcode="gycon:89109">
  <name>Las Vegas</name>
  <broadener qcode="iso3166-2:US-NV">
    <name>Nevada</name>
  </broadener>
  <broadener qcode="iso3166-1a3:USA">
    <name>United States</name>
  </broadener>
</subject>
<keyword>business</keyword>
<keyword>economic</keyword>
<keyword>economy</keyword>
<keyword>finance</keyword>
<keyword>poor</keyword>
<keyword>poverty</keyword>
<keyword>gamble</keyword>
<headline>Variety Of Recessionary Forces Leave Las Vegas
  Economy Scarred</headline>
<description role="drol:caption">A general view of part of downtown,
  including Las Vegas Boulevard, on October 20, 2010 in Las Vegas,
  Nevada. Nevada once had among the lowest unemployment rates in the
  United States at 3.8 percent but has since fallen on difficult times.
  Las Vegas, the gaming capital of America, has been especially hard
  hit with unemployment currently at 14.7 percent and the highest
  foreclosure rate in the nation. Among the sparkling hotels and
  casinos downtown are dozens of dormant construction projects and
  hotels offering rock bottom rates. As the rest of the country slowly
  begins to see some economic progress, Las Vegas is still in the midst
  of the economic downturn. (Photo by Spencer Platt/Getty Images)
</description>
</contentMeta>

```

9.7 Picture data

Binary content is conveyed within the NewsML-G2 <contentSet> wrapper by one or more <remoteContent> elements, enabling multiple alternate renditions of a picture within the same Item.

9.7.1 Remote Content

The <remoteContent> element references objects that exist independently of the current NewsML-G2 Item. In the example there is an instance of <remoteContent> for each of the three separate binary renditions of the picture.

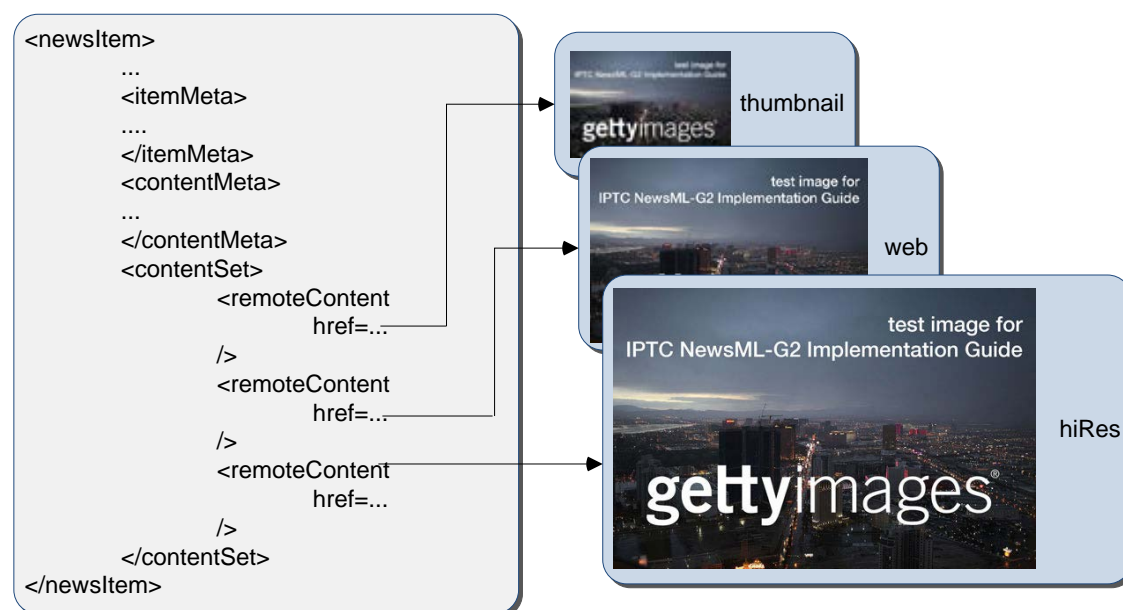


Figure 5: Each `<remoteContent>` wrapper references a separate rendition of the binary resource

Each remote content instance contains attributes that conceptually can be split into three groups:

- ❖ **Target resource attributes** enable the receiver to accurately identify the remote resource, it's content type and size;
- ❖ **Content attributes** enable the processor to distinguish the different business purposes of the content using `@rendition`;
- ❖ **Content Characteristics** contain technical metadata such as dimensions, colour values and resolution.

Frequently used attributes from these groups are described below, but note that the NewsML-G2 XML structure that delimits the groups may not be visible in all XML editors. For a detailed description of these attribute groups, see the NewsML-G2 Specification (if not at hand, this can be downloaded by visiting www.newsml-g2.org and navigating to Specification resources).

9.7.2 Target Resource Attributes

This group of attributes express administrative metadata, such as identification and versioning, for the referenced content, which could be a file on a mounted file system, a Web resource, or an object within a content management system. G2 flexibly supports alternative methods of identifying and locating the externally-stored content. For this example, the picture renditions are located in the same folder as the NewsML-G2 document.

The two attributes of `<remoteContent>` available to identify and locate the content are Hyperlink (`@href`) and Resource Identifier Reference (`@residref`). Either one **MUST** be used to identify and locate the target resource. They **MAY** optionally be used together. Although these two attributes are superficially similar, their intended use is:

- ❖ `@href` *locates* any resource, using an IRI.
- ❖ `@residref` *identifies* a *managed* resource, using an identifier that may be globally unique.

9.7.2.1 Hyperlink (`@href`)

An IRI, for example:

```
<remoteContent href="http://example.com/2008-12-20/pictures/foo.jpg"
```

Or (amongst other possibilities):

```
<remoteContent href="file:///./GYI0062134533-web.jpg"
```

9.7.2.2 Resource Identifier Reference (`@residref`)

An XML Schema string, such as:



```
<remoteContent residref="tag:example.com,2008:PIX:F0020081220098658"
```

It is up to the provider to specify how @residref may be resolved to retrieve the actual content.

9.7.2.3 Version

An XML Schema positive integer denoting the version of the target resource. In the absence of this attribute, recipients should assume that the target is the latest available version:

```
<remoteContent href="file:/// / GYI0062134533-web.jpg" version="1"
```

9.7.2.4 Content Type

The MIME type of the target resource:

```
contenttype="image/jpeg"
```

9.7.2.5 Size

Indicates the size of the target resource in bytes.

```
size="346071"
```

9.7.3 News Content Attributes

This group of attributes of <remoteContent> enables a processor or human-reader to distinguish between different components; in this case the alternative resolutions of the picture. The principal attribute of this group is @rendition, described below.

9.7.3.1 Rendition

The rendition attribute MUST use a QCode, either proprietary or using the IPTC NewsCode for rendition, which has a Scheme URI of <http://cv.iptc.org/newscodes/rendition/> and recommended Scheme Alias of "rnd" and contains (amongst others) the values that we need: highRes, web, thumbnail. Thus using the appropriate NewsCode, the high resolution rendition of the picture may be identified as:

```
<remoteContent rendition="rnd:highRes"
```

To avoid processing ambiguity, each specific rendition value should be used only once per News Item, except when the same rendition is available from multiple remote locations. In this case, the same value of rendition may be given to several Remote Content elements.

9.7.4 News Content Characteristics

This group of attributes describes the physical properties of the referenced object specific its media type. Text, for example, may use @wordcount). Audio and video are provided with attributes appropriate to streamed media, such as @audiobitrate, @videoframerate. The appropriate attributes for pictures are described below.

9.7.4.1 Picture Width and Picture Height

The dimension attributes @width and @height are optionally qualified by @dimensionunit, which specifies the units being used. This is a @qcode value and it is recommended that the value is taken from the IPTC Dimension Unit NewsCodes, whose URI is <http://cv.iptc.org/newscodes/dimensionunit/> (recommended Scheme Alias is "dimensionunit")

If @dimensionunit is absent, the default units for each content type are:

Content Type	Height Unit (default)	Width Unit (default)
Picture	pixels	pixels
Graphic: Still / Animated	points	points
Video (Analog)	lines	pixels

Content Type	Height Unit (default)	Width Unit (default)
Video (Digital)	pixels	pixels

The dimensions of the example picture are expressed in pixels, the default, therefore @dimensionunit is not needed:

`width="480"`
`height="2075"`

9.7.4.2 Picture Orientation

This indicates that the image requires an orientation change before it can be properly viewed, using values of 1 to 8 (inclusive), where 1 (the default) is "upright": that is the visual top of the picture is at the top, and the visual left side of the picture is on the left.

The application of these orientation values is described in detail in the *News Content Characteristics* section of the NewsML-G2 Specification. (If not at hand, this can be downloaded by visiting www.newsml-g2.org and navigating to Specification resources.)

The example picture above has an orientation value of 1:

`width="1500"`
`height="1001"`
`orientation="1"`

9.7.4.3 Layout Orientation

It is possible to calculate the best way to use a picture in a page layout using the combined technical characteristics of Height, Width and Orientation, but many implementers are reluctant to rely on technical characteristics to make editorial judgements (determining whether a video is SD or HD is another example). The @layoutorientation is a way to express editorial advice on the best way to use a picture in a layout. The value for the example picture is:

`layoutorientation="layout: horizontal"`

Values in the Layout Orientation Scheme are:

Code	Definition
horizontal	The human interpretation of the top of the image is aligned to the long side.
vertical	The human interpretation of the top of the image is aligned to the short side.
Square	Both sides of the image are about identical, there is no short and long side.
unaligned	There is no human interpretation of the top of the image.

9.7.4.4 Picture Colour Space

The colour space of the target resource, and MUST use a QCode. The recommended scheme is the IPTC Colour Space NewsCodes (recommended scheme alias "colsp") Note the UK English spelling of colour.

`colourspace="colsp: AdobeRGB"`

9.7.4.5 Colour Depth

The optional @colourdepth indicates using a non-negative integer the number of bits used to define the colour of each pixel in a still image, graphic or video.

`colourdepth="24"`

9.7.4.6 Content Hints

At the Power conformance level, the provider is able to express metadata from the target resource³ as an aid to processing. In this case, the provider has added an <altId> – an alternative identifier – for the resource.

Alternative identifiers may be needed by customer systems. The <altId> element may optionally be refined using a QCode to describe the context – in this case a “master ID” that is proprietary to the provider. This makes clear the purpose of the alternative identifier. Also note that Alternative Identifiers are useful only to another application; and not intended to be used by THIS NewsML-G2 processor. The provider MUST tell receivers how to interpret alternative identifiers, otherwise they are meaningless.

```
<altId type="gyi id: masterID">105864332</altId>
```

Note that in this example only the high resolution rendition has an <altId>.

9.7.4.7 Signal

The signal property instructs the G2 processor to process an Item or its content in a specific way. As a child element of itemMeta, the scope of <signal> is the whole of the document and/or its contents. If alternative renditions of content have specific processing needs, use <signal> as a child element of <remoteContent> to specify the processing instructions.

9.7.5 Completed <remoteContent> wrapper

The <remoteContent> wrapping element in full for the “High Res” picture in the example:

```
<remoteContent rendition="rnd: highRes"
  href=". /GYI0062134533.jpg" version="1"
  size="346071" contentType="image/jpeg" width="1500"
  height="1001" colourspace="colsp: AdobeRGB" orientation="1"
  layoutorientation="layoutorient: horizontal">
  <altId type="gyi id: masterID">105864332</altId>
</remoteContent>
```

9.8 The Complete Picture

The Listing below is given for completeness. The pictures are identified/located using @href.

LISTING 3 Photo in NewsML-G2

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:gettyimages.com, 2010: GYI0062134533"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef href="http://cv.gettyimages.com/nml2catalog4customers-1.xml" />
  <rightsInfo>
    <copyrightHolder uri="http://www.gettyimages.com">
      <name>Getty Images North America</name>
    </copyrightHolder>
    <copyrightNotice
      href="http://www.gettyimages.com/Corporate/LicenseInfo.aspx">
      (c) 2010 Copyright Getty Images. --
      http://www.gettyimages.com/Corporate/LicenseInfo.aspx
    </copyrightNotice>
    <usageTerms>Contact your local office for all commercial or
      promotional uses. Full editorial rights UK, US, Ireland, Canada (not
      Quebec). Restricted editorial rights for daily newspapers elsewhere,
      please call.</usageTerms>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ni nat: picture"/>
    <provider qcode="nprov: GYI">
```

³ It is not mandatory for the metadata to be extracted from the target resource, but it MUST agree with any existing metadata within the target resource.


```

    <name>Getty Images Inc.</name>
  </provider>
  <versionCreated>2014-11-12T06:42:04Z</versionCreated>
  <firstCreated>2010-10-20T20:58:00Z</firstCreated>
</itemMeta>
<contentMeta>
  <contentCreated>2010-10-20T19:45:58-04:00</contentCreated>
  <creator role="crol:photographer">
    <name>Spencer Platt</name>
    <related rel="crel:isa" qcode="gyibt:staff" />
  </creator>
  <contributor role="ctrol:descrWriter">
    <name>sp/lrc</name>
  </contributor>
  <creditline>Getty Images</creditline>
  <subject type="cpnat:event" qcode="gyimeid:104530187" />
  <subject type="cpnat:abstract" qcode="medtop:20000523">
    <name xml:lang="en-GB">labour market</name>
    <name xml:lang="de">Arbeitsmarkt</name>
  </subject>
  <subject type="cpnat:abstract" qcode="medtop:20000533">
    <name xml:lang="en-GB">unemployment</name>
    <name xml:lang="de">Arbeitslosigkeit</name>
  </subject>
  <subject type="cpnat:geoArea">
    <name>Las Vegas Boulevard</name>
  </subject>
  <subject type="cpnat:geoArea" qcode="gycon:89109">
    <name>Las Vegas</name>
    <broadener qcode="iso3166-2:US-NV">
      <name>Nevada</name>
    </broadener>
    <broadener qcode="iso3166-1a3:USA">
      <name>United States</name>
    </broadener>
  </subject>
  <keyword>business</keyword>
  <keyword>economic</keyword>
  <keyword>economy</keyword>
  <keyword>finance</keyword>
  <keyword>poor</keyword>
  <keyword>poverty</keyword>
  <keyword>gamble</keyword>
  <headline>Variety Of Recessionary Forces Leave Las Vegas
    Economy Scarred</headline>
  <description role="drol:caption">A general view of part of downtown, including
    Las Vegas Boulevard, on October 20, 2010 in Las Vegas, Nevada. Nevada once
    had among the lowest unemployment rates in the United States at 3.8 percent
    but has since fallen on difficult times. Las Vegas, has been especially hard
    hit with unemployment currently at 14.7 percent. Among the sparkling hotels
    and casinos downtown are dozens of dormant construction projects and hotels
    offering rock bottom rates. As the rest of the country slowly begins to see
    some economic progress, Las Vegas is still in the midst of the economic
    downturn. (Photo by Spencer Platt/Getty Images)
  </description>
</contentMeta>
<contentSet>
  <remoteContent rendition="rnd:highRes">
    href=". /GYI0062134533.jpg" version="1"
    size="346071" contenttype="image/jpeg" width="1500"
    height="1001" colourspace="colsp:AdobeRGB" orientation="1"
    layoutorientation="loutorient:horizontal">
    <altId type="gyiid:masterID">105864332</altId>
  </remoteContent>
  <remoteContent rendition="rnd:web">
    href="file:/// / GYI0062134533-web.jpg" version="1"
    size="28972" contenttype="image/jpeg" width="480"
    height="320" colourspace="colsp:AdobeRGB" orientation="1"
    layoutorientation="loutorient:horizontal">
  </remoteContent>
  <remoteContent rendition="rnd:thumb">
    href="file:/// / GYI0062134533-thumb.gif" version="1"
    size="6381" contenttype="image/gif" width="80"
    height="53" colourspace="colsp:AdobeRGB" orientation="1"
    layoutorientation="loutorient:horizontal">
  </remoteContent>
</contentSet>
</newsItem>

```



This page intentionally blank

10 Quick Start: Video

10.1 Introduction

Now that streamed media is part of everyone's day-to-day experience on the Web, organisations with little or no tradition of "broadcast media" production need to be able to process audio and video.

NewsML-G2 allows all media organisations, whether traditional broadcasters or not, to access and exchange audio and video in a professional workflow, by providing features and Extension Points that enable proprietary formats to be "mapped" to Newsml-G2 to achieve freedom of exchange amongst a wider circle of information partners.

This Quick Start guide is split into two parts:

- ❖ Part I deals with a video that is available in multiple different renditions and the example focuses on expressing the technical characteristics of each rendition of the content.
- ❖ Part II shows an example of video content has been assembled from multiple sources, each with distinct metadata.



Read the Quick Guide to G2 Basics
before this Quick Start.

10.2 Part I – Multiple Renditions of a Single Video

The following example is based on a sample NewsML-G2 video item from Agence France Presse (but is not a guide to processing AFP's G2 news services).

10.2.1 Document structure

The building blocks of the G2 Item are the <newsItem> root element, with additional wrapping elements for metadata about the News Item (itemMeta), metadata about the content (contentMeta) and the content itself (contentSet).

The root <newsItem> attributes are:

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"  
  guid="urn:newsml:afp.com:20140131:ENG.3424d3807bc.391@video_1359566"  
  version="3"  
  standard="NewsML-G2"  
  standardversion="2.18"  
  conformance="power"  
  xml:lang="en-US">
```

This is followed by Catalog references:

```
<catalogRef  
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />  
<catalogRef href="http://cv.afp.com/std/catalog/catalog.AFP-IPTC-G2_3.xml" />
```

10.2.2 Item Metadata <itemMeta>

The <itemClass> property uses a QCode from the IPTC News Item Nature NewsCodes to denote that the Item conveys a picture. Note that <provider> uses the recommended IPTC Provider NewsCodes, a controlled vocabulary of providers registered with the IPTC, recommended scheme alias "nprov":

```
<itemMeta>  
  <itemClass qcode="ni:nat:video" />  
  <provider qcode="nprov:AFP" />  
  <versionCreated>2014-01-31T11:37:23+01:00</versionCreated>  
  <firstCreated>2014-01-30T13:29:38+00:00</firstCreated>  
  <pubStatus qcode="stat:usable" />  
</itemMeta>
```

10.2.3 Content Metadata <contentMeta>

The <icon> element tells receivers how to retrieve an image to use as an iconic image for the content, for example a still image extracted from the video. It's possible to have multiple icons to suit different applications, each qualified by @rendition.

Two <description> elements are qualified by @role: first a summary, second a more detailed shotlist:

```
<contentMeta>
  <icon contentType="image/jpeg" height="62"
    href="http://spar-iris-p-sco-http-int-vip.afp.com/components/9601ac3"
    rendition="rnd:thumbnail" width="110" />
  <creditline>AFP</creditline>
  <description role="afpdescRole:synthe">- Amir Hussein Abdullahian, Iranian
    foreign ministry's undersecretary for Arab and African affairs -
    Panos Mountzis (man), UNHCR regional coordinator for Syrian refugees
  </description>
  <description role="afpdescRole:script">SHOTLIST: KUWAIT. JANUARY 30,
    2014. SOURCE: AFPTV -VAR inside the conference room -VAR of Ban
    Ki-moon -MS of King Abdullah II of Jordan -MS of Michel Sleiman,
    president of Lebanon -MS of Tunisian president Moncef Marzouki
    SOUNDBITE 1 - Amir Hussein Abdullahian (man), Iranian foreign
    ministry's undersecretary for Arab and African affairs (Farsi, 10
    sec): "Those who send arms to Syria are behind the daily killings
    there." SOUNDBITE 2 - Amir Hussein Abdullahian (man), Iranian foreign
    ministry's undersecretary for Arab and African affairs (Farsi, 9
    sec): "We regret that some countries, such as the United States, have
    created a very high level of extremism in Syria." SOUNDBITE 3 - Panos
    Mountzis (man), UNHCR regional coordinator for Syrian refugees
    (Arabic, 12 sec): "The United Nations is providing humanitarian
    assistance to more than four million people inside Syria, two million
    of them displaced." SOUNDBITE 4 - Panos Mountzis (man), UNHCR
    regional coordinator for Syrian refugees (Arabic, 17 sec): "The
    funding will first go to UN relief organizations, who are working
    inside Syria and in neighbouring countries. Funding will also go to
    the more than 55 NGOs in Syria with whom we cooperate and coordinate
    to deliver aid."
  </description>
  <language tag="en" />
</contentMeta>
```

10.2.4 Video Content

Video is conveyed within the NewsML-G2 <contentSet> using the <remoteContent> element; where there are multiple alternate renditions of SAME content, <remoteContent> can be repeated for each rendition within the same Item.

The <remoteContent> element references binary objects that exist independently of the current NewsML-G2 document. In this example there is an instance of <remoteContent> for each of three renditions of the video.

Each remote content instance contains attributes that can conceptually be split into three groups:

- ❖ **Target resource attributes** enable the receiver to accurately identify the remote resource, its content type and size;
- ❖ **Content attributes** enable the processor to distinguish the different business purposes of the content using @rendition;
- ❖ **Content Characteristics** contain technical metadata such as dimensions, duration and format.

Frequently used attributes from these groups are described below, but note that the NewsML-G2 XML structure that delimits the groups may not be visible in all XML editors. For a detailed description of these attribute groups, see the NewsML-G2 Specification (If not at hand, this can be downloaded by visiting www.newsml-g2.org and navigating to Specification resources.)

10.2.5 Target Resource Attributes

This group of attributes express administrative metadata, such as identification and versioning, for the referenced content, which could be a file on a mounted file system, a Web resource, or an object within a content management system. G2 flexibly supports alternative methods of identifying and locating the externally-stored content.

The two attributes of <remoteContent> that identify and optionally locate the content are Hyperlink (@href) and Resource Identifier Reference (@residref). Either one MUST be used to identify the target resource. They MAY optionally be used together.

Although @href and @residref are superficially similar, their intended use is:

- ❖ @href *locates* any resource, using an IRI.
- ❖ @residref *identifies* a *managed* resource, using an identifier that may be globally unique. It is up to the provider to specify how @residref may be resolved to retrieve the actual content.

10.2.5.1 Hyperlink (@href)

An IRI, for example:

```
<remoteContent  
  href="http://components.afp.com/ab652af034e5f7acc131f8f122b274a5ef8ee37e.mpg"
```

10.2.5.2 Resource Identifier Reference (@residref)

An XML Schema string e.g.

```
<remoteContent residref="tag:example.com, 2008:PIX:F0020081220098658"
```

10.2.5.3 Version

An XML Schema positive integer denoting the version of the target resource. In the absence of this attribute, recipients should assume that the target is the latest available version

```
  version="1"
```

10.2.5.4 Content Type

The MIME type of the target resource

```
  contenttype="video/3gpp"
```

10.2.5.5 Format

A refinement of a Content Type using a value from a controlled vocabulary:

```
  format="fmt:video"
```

10.2.5.6 Content Type Variant (@contenttypevariant)

A refinement of a Content Type using a string:

```
  contenttype="video/3gpp"  
  contenttypevariant="MPEG-4 Simple Profile"
```

10.2.5.7 Size

Indicates the size of the target resource in bytes.

```
  size="54593540"
```

10.2.6 News Content Attributes

This group of attributes of <remoteContent> enables a processor or human operator to distinguish between different components; in this case the alternative resolutions of the video.

10.2.6.1 Rendition

The rendition attribute MUST use a QCode. Providers may have their own schemes, or use the IPTC NewsCode for rendition, which has a Scheme URI of <http://cv.iptc.org/newscodes/rendition/> and recommended Scheme Alias of "rnd". This example uses a provider-specific scheme with a Scheme Alias of "vidrnd":



`<remoteContent rendition="vi drnd: dvd">`

To avoid processing ambiguity, each specific rendition value should be used only once per News Item, except when the same rendition is available from multiple remote locations. In this case, the same value of rendition may be given to several Remote Content elements.

10.2.7 News Content Characteristics

This third a group of attributes of `<remoteContent>` is provided to enable further efficiencies in processing and describes physical characteristics of the referenced object specific its media type. Text, for example, may use `@wordcount`; Audio and video are provided with attributes appropriate to streamed media, such as `@audiobitrate`, `@videoframerate`. The appropriate attributes for video are described below.

10.2.7.1 Duration (@duration and @durationunit)

Indicates the duration of the content in seconds by default, but can be expressed by some other measure of temporal reference (e.g. frames) when using the optional `@durationunit`. From NewsML-G2 2.14, the data-type of `@duration` is a string; earlier versions use non-negative integer. The reason for the change is that video duration is often expressed using non-integer values.

For example, expressing duration as an SMPTE time code requires the following NewsML-G2:

`duration="00: 06: 32: 12" durationunit="timeunit: timeCode"`

The recommended CV for `@durationunit` is the IPTC Time Unit NewsCodes whose URI is <http://cv.iptc.org/newscodes/timeunit/>. The recommended alias for the scheme is "timeunit".

10.2.7.2 Video Codec (@videocodec)

A QCode value indicating the encoding of the video – for example one of the encodings used in this example is MPEG-2 Video Simple Profile. This is indicated by the IPTC Video Codec NewsCodes with a recommended Scheme Alias "vcdc", and the corresponding code is "c015".

`videocodec="vcdc: c015"`

10.2.7.3 Video Frame Rate (@videoframerate)

A decimal value indicating the rate, in frames per second [fps] at which the video should be played out in order to achieve the correct visual effect. Common values (in fps) are 25, 50, 60 and 29.97 (drop-frame rate):

`videoframerate="25"`

10.2.7.4 Video Aspect Ratio (@videoaspectratio)

A string value, e.g. 4:3, 16:9

`videoaspectratio="4: 3"`

10.2.7.5 Video Scaling (@videoscaling)

The `@videoscaling` attribute describes how the aspect ratio of a video has been changed from the original in order to accommodate a different display dimension:

`videoscaling="sov: letterboxed"`

The value of the property is a QCode; the recommended CV is the IPTC Video Scaling NewsCode (Scheme URI : <http://cv.iptc.org/newscodes/videoscaling/>)

The recommended Scheme Alias is "sov", and the codes and their definitions are as follows:

Code	Definition
unscaled	no scaling applied

Code	Definition
mixed	two or more different aspect ratios are used in the video over the timeline
pillarboxed	bars to the left and right
letterboxed	bars to the top and bottom
windowboxed	pillar boxed plus letter boxed
zoomed	scaling to avoid any borders

10.2.7.6 Video Definition (@videodefinition)

Editors may need to know whether video content is HD or SD, as this may not be obvious from the technical specification ("HD", for example, is an umbrella term covering many different sets of technical characteristics). The @videodefinition attribute carries this information:

videodefinition="videodef:sd"

The value of the property can be either "hd" or "sd", as defined by the Video Definition NewsCodes CV. The Scheme URI is <http://cv.iptc.org/newscodes/videodefinition/> and the recommended scheme alias is "videodef".

10.2.7.7 Colour Indicator <colourindicator>

Indicates whether the still or moving image is coloured or black and white (note the UK spelling of colour). The recommended vocabulary is the IPTC Colour Indicator NewsCodes (Scheme URI: <http://cv.iptc.org/newscodes/colourindicator/>) with a recommended Scheme Alias of "colin". The value of the property is "bw" or "colour":

colourindicator="colin:colour"

The completed Remote Content wrapper will be:

```
<remoteContent contentType="video/mpeg-2"
href="http://components.afp.com/ab652af034e5f7acc131f8f122b274a5ef8ee37e.mpg"
rendition="vdrnd:dvd"
size="54593540"
width="720" height="576"
duration="69"
durationunit="timeunit:seconds"
videocodec="vdc:c015"
videoframerate="25"
videodefinition="videodef:sd"
colourindicator="colin:colour"
videoaspectratio="4:3"
videoscaling="sov:letterboxed" />
```

10.3 Audio metadata

There are specific properties for describing the technical characteristics of audio, for example:

10.3.1.1 Audio Bit Rate (@audiobitrate)

A positive integer indicating kilobits per second (Kbps)

audiobitrate="32"

10.3.1.2 Audio Sample Rate (@audiosamplerate)

A positive integer indicating the sample rate in Hertz (Hz)

audiosamplerate="44100"

For a detailed description of all of the News Content Characteristics for Video and Audio content, see section 13.9.9 News Content Characteristics in the NewsML-G2 Specification Document



10.4 Full Part 1 listing

LISTING 4 Multiple Renditions of a Video in NewsML-G2

```
<?xml version="1.0" encoding="utf-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:afp.com:20140131:CNG.3424d3807bc.391@video_1359566"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef href="http://cv.afp.com/std/catalog/catalog.AFP-IPTC-G2_3.xml" />
  <itemMeta>
    <itemClass qcode="ni nat: video" />
    <provider qcode="nprov: AFP" />
    <versionCreated>2014-01-31T11:37:23+01:00</versionCreated>
    <firstCreated>2014-01-30T13:29:38+00:00</firstCreated>
    <pubStatus qcode="stat: usable" />
  </itemMeta>
  <contentMeta>
    <icon contenttype="image/jpeg" height="62"
      href="http://spar-iris-p-sco-http-int-vip.afp.com/components/9601ac3"
      rendition="rnd: thumbnail" width="110" />
    <creditline>AFP</creditline>
    <description role="afpdescRole: synthe">- Amir Hussein Abdullahian, Iranian
      foreign ministry's undersecretary for Arab and African affairs -
      Panos Mountzis (man), UNHCR regional coordinator for Syrian refugees
    </description>
    <description role="afpdescRole: script">SHOTLIST: KUWAIT. JANUARY 30,
      2014. SOURCE: AFPTV -VAR inside the conference room -VAR of Ban
      Ki-moon -MS of King Abdullah II of Jordan -MS of Michel Sleiman,
      president of Lebanon -MS of Tunisian president Moncef Marzouki
      SOUNDBITE 1 - Amir Hussein Abdullahian (man), Iranian foreign
      ministry's undersecretary for Arab and African affairs (Farsi, 10
      sec): "Those who send arms to Syria are behind the daily killings
      there." SOUNDBITE 2 - Amir Hussein Abdullahian (man), Iranian foreign
      ministry's undersecretary for Arab and African affairs (Farsi, 9
      sec): "We regret that some countries, such as the United States, have
      created a very high level of extremism in Syria." SOUNDBITE 3 - Panos
      Mountzis (man), UNHCR regional coordinator for Syrian refugees
      (Arabic, 12 sec): "The United Nations is providing humanitarian
      assistance to more than four million people inside Syria, two million
      of them displaced." SOUNDBITE 4 - Panos Mountzis (man), UNHCR
      regional coordinator for Syrian refugees (Arabic, 17 sec): "The
      funding will first go to UN relief organizations, who are working
      inside Syria and in neighbouring countries. Funding will also go to
      the more than 55 NGOs in Syria with whom we cooperate and coordinate
      to deliver aid."
    </description>
    <language tag="en" />
  </contentMeta>
  <contentSet>
    <remoteContent contenttype="video/mpeg-2"
      href="http://components.afp.com/ab652af034e.mpg"
      rendition="vi drnd: dvd"
      size="54593540"
      width="720" height="576"
      duration="69"
      durationunit="timeunit: seconds"
      videocodec="vcdc: c015"
      videoframerate="25"
      videodefinition="videodef: sd"
      colourindicator="col in: colour"
      videoaspectratio="4: 3"
      videoscaling="sov: letterboxed" />
    <remoteContent contenttype="video/mp4-1920x1080"
      href="http://components.afp.com/3e353716caa.1920x1080.mp4"
      rendition="vi drnd: HD1080"
      size="87591736"
      width="1920" height="1080"
      duration="69"
      durationunit="timeunit: seconds"
      videocodec="vcdc: c041"
      videoframerate="25"
      videodefinition="videodef: hd"
    </remoteContent>
  </contentSet>
</newsItem>
```



```
    colourindicator="col in: colour"  
    videoaspectratio="16: 9"  
    videoencoding="sov: unscal ed" />  
  <remoteContent contentType="video/mp4- 1280x720"  
    href="http://components. afp. com/5ba0d14a64f. 1280x720. mp4"  
    rendition="vi drnd: HD720"  
    size="71010540"  
    width="1280" height="720"  
    duration="69"  
    durationunit="timeunit: seconds"  
    videocodec="vcdc: c041"  
    videoframerate="25"  
    videodefinition="vi deodef: hd"  
    colourindicator="col in: colour"  
    videoaspectratio="16: 9"  
    videoencoding="sov: unscal ed" />  
</contentSet>  
</newsItem>
```

10.5 Part 2 – Multi-part video



Read the Quick Start guide “G2 Basics” and the preceding Part 1 of this guide to video before reading Part 2.

10.6 Introduction

Audio and video, including animation, have a temporal dimension: the nature of the content is expected to change over its duration: in this example a single piece of video has been created from a number of shots – shorter segments of content – from several different creators that were combined during an editing process.⁴ Each segment of streamed content has its own metadata, in addition to the metadata that applies to the content as a whole.

In addition to metadata structures that apply to the whole content, NewsML-G2 can express metadata about separate identifiable parts of content using <partMeta>.

The example video is about a retrospective exhibition in Berlin of works by the German humorist and animator Vicco von Bülow. It consists of a number of shots, so provides a shotlist summarising the visual content of each shot, and a dopesheet, giving an editorial summary of the video's content.

The document structure and the G2 properties included in the example have been previously described, except for the <partMeta> wrapper, which is described in detail below. A full code listing for the example is included at the end.

The example is based on a sample NewsML-G2 video item from the European Broadcasting Union (EBU). The News Item references a multi-part broadcast video and contains separate metadata for each segment of the content, including a keyframe, and additionally describes the overall technical characteristics of the video.

Please note that it may resemble but does NOT represent the EBU's NewsML-G2 implementation.

10.6.1 Part Metadata

G2 Items can have many <partMeta> wrappers, each expressing properties for an identifiably separate part of the content; in this example each of the shots, or segments, which make up the video. The properties for each segment include:

- ❖ an ID for the segment, and a sequence number
- ❖ a keyframe, or icon that may help to visually identify the content of the segment
- ❖ the start time and duration of the segment

⁴ Note that this complies with the basic NewsML-G2 rule that “one piece of content = one newsItem”. Although the video may be composed of material from many sources, it remains a single piece of journalistic content created by the video editor. This is analogous to a text story that is compiled by a single reporter or editor from several different reports.

It is also possible to assert any Administrative or Descriptive Metadata for each <partMeta> element, if required.

The id and sequence number for the shot are expressed as attributes of <partMeta>:

```
<partMeta parti d="Part1_ID" seq="1">
```

10.6.1.1 Add keyframe using <icon>

A keyframe for the video segment is expressed as the child element <icon> with @href pointing to the keyframe image as a resource on the Web:

```
<i con href="http://www.example.com/video/Keyframes/407624.jpeg"/>
```

10.6.1.2 Timing metadata

The <timeDelim> property indicates the start and end time of this segment, and the time unit being used as shown below:

```
<timeDelim start="0" end="446" timeunit="timeunit:editUnit"/>
```

In the example, @start and @end are expressed as integers but note that the datatype of @start and @end is XML String, because start and end times can be expressed as integer, time, or SMPTE time code.

@timeunit use a QCode to indicate that @start and @end are expressed as frames, the smallest editable unit of video content. This is the default and is assumed if @timeunit is not present in the <timeDelim> element. Other values of time unit MUST be taken from the IPTC Time Unit NewsCodes (recommended Scheme Alias "timeunit").

The values in the scheme are:

- ❖ editUnit: (default) the time is expressed in frames (video) or samples (audio).
- ❖ timeCode: the format of the timestamp is hh:mm:ss:ff (ff for frames).
- ❖ timeCodeDropFrame: the format of the timestamp is hh:mm:ss:ff (ff for frames).
- ❖ normalPlayTime: the format of the timestamp is hh:mm:ss.sss (milliseconds).
- ❖ seconds.
- ❖ milliseconds.



When specifying the start and end points of a video segment, be aware that these are unlikely to be frame-accurate for the same shot rendered in different technical formats, due to potentially differing frame-rates. They may even produce different viewing results.

To overcome this issue, the attribute @renditionref has been added to <timeDelim>, allowing implementers to specify separate time delimiters corresponding to alternative renditions of the same shot. For example:

```
<partMeta...>
  <!-- 10 seconds in frames at 25 fps = 250 frames -->
  <timeDelim renditionref="vi drnd: avi 25" start="0" end="250"
    timeunit="timeunit:editUnit"/>
  <!-- 10 seconds in frames at 30 fps = 300 frames -->
  <timeDelim renditionref="vi drnd: avi 30" start="0" end="300"
    timeunit="timeunit:editUnit"/>
</partMeta>
```

Each @renditionref identifies the corresponding @rendition in <remoteContent>:

```
<contentSet>
  <remoteContent contenttype="..."
    href="..."
    rendition="vi drnd: avi 25"
  ... />
  <remoteContent contenttype="..."
    href="..."
    rendition="vi drnd: avi 30"
  ... />
</contentSet>
```




10.6.1.3 Description and Language

The example also indicates the language being used in the shot, and the context in which it is used. In this case, @role uses a QCode from a proprietary EBU scheme to indicate that the soundtrack of the shot is a voiceover in English.

```
<language tag="en" role="langusecode: Voi ce0ver" />
```

Implementers may also use the IPTC Language Role NewsCodes (recommended Scheme Alias "lrol") for this purpose.

Using <description>, we can also indicate what the viewer can expect to see in this segment:

```
<description>Vico von Bülow entering exhibition </description>
```

10.6.1.4 <partMeta> example

The <partMeta> element is repeated for each of the video segments. Below is a single complete <partMeta> from the example:

```
<partMeta partid="Part1_ID" seq="1">
  <icon href="http://www.example.com/video/Keyframes/407624.jpeg"/>
  <timeDelim start="0" end="446" timeunit="timeunit: editUnit"/>
  <language tag="en" role="lrol: voi ce0ver" />
  <description>Vico von Bülow entering exhibition </description>
</partMeta>
```

10.6.2 Video Content

The <contentSet> wrapper contains a single rendition of the video inside the <remoteContent> element:

```
<contentSet>
  <remoteContent href="http://www.example.com/video/407624.avi"
    format="fmt: avi"
    duration="111" durationunit="timeunit: seconds"
    videocodec="vcdc: c155"
    videoframerate="25"
    videoaspectratio="16: 9" />
</contentSet>
```

10.7 Full Part 2 listing

LISTING 5 Multi-part Video in NewsML-G2

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:example.com,2008:407624"
  version="4"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef
    href="http://www.example.com/metadata/newsml-g2/catalog.NewsML-G2.xml" />
  <rightsInfo>
    <usageTerms>
      Access only for Eurovision Members and EVN / EVS Sub-Licensees.
    </usageTerms>
    <usageTerms>
      Coverage cannot be used by a national competitor of the contributing
      broadcaster.
    </usageTerms>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ni nat: video" />
    <provider qcode="providercode: EBU">
      <name>European Broadcasting Union - EVN</name>
      <organizationDetails>
        <contactInfo>
          <web>http://www.eurovision.net</web>
          <phone>+41 22 717 2869</phone>
          <email>features@eurovision.net</email>
          <address role="AddressType: Office">
            <line>Eurovision Sports News Exchanges</line>
          </address>
        </contactInfo>
      </organizationDetails>
    </provider>
  </itemMeta>
</newsItem>
```

```

        <line>L Ancienne Route 17 A</line>
        <line>CH- 1218</line>
        <locality>
            <name>Grand- Saconnex</name>
        </locality>
        <country qcode="ISOCountryCode: ch">
            <name xml:lang="en">Swi tzerland</name>
        </country>
    </address>
</contactInfo>
</organisationDetails>
</provider>
<versionCreated>2012- 11- 08T10: 54: 04Z</versionCreated>
<firstCreated>2008- 11- 05T15: 22: 28Z</firstCreated>
<pubStatus qcode="stat: usable" />
<service qcode="servi cecode: EUROVI SION">
    <name>Eurovi si on servi ces</name>
</service>
<edNote>Originally broadcast in Germany</edNote>
<link rel="irel: associatedWith"
    href="http://www. exampl e. com/vi deo/407624/i index. html "/>
</itemMeta>
<contentMeta>
    <contentCreated> 2008- 12- 22T23: 04: 00- 08: 00</contentCreated>
    <located type="cptype: ci ty" qcode="ci ty: 345678">
        <name>Berli n</name>
        <broadr type="cptype: statprov" qcode="state: 2365">
            <name>Berli n</name>
        </broadr>
        <broadr type="cptype: country" qcode="i so3155- i a2: DE">
            <name>Germany</name>
        </broadr>
    </located>
    <creator qcode="codesource: DEZDF">
        <name>Zweites Deutsches Fernsehen</name>
        <organisationDetails>
            <location>
                <name>MAI NZ</name>
            </location>
        </organisationDetails>
    </creator>
    <contributor qcode="codeorigi n: DEZDF" rol e="rol ecode: Techni cal Ori gi n">
        <name>Zweites Deutsches Fernsehen</name>
    </contributor>
    <creator qcode="codesource: GBRTV">
        <name>Reuters Televisi on Ltd</name>
    </creator>
    <language tag="en" rol e="langusecode: Voi ce0ver" >
        <name>Engli sh</name>
    </language>
    <genre qcode="genre: bi og">
        <name xml:lang="en- GB">Bi ography</name>
        <name xml:lang="fr">bi ographi e</name>
    </genre>
    <subject type="cpnat: abstract" qcode="medtop: 01000000">
        <name xml:lang="en- GB">Arts, Culture and Entertainment</name>
        <name xml:lang="fr">Arts, culture, et spectacles</name>
        <narrower type="cpnat: abstract" qcode="medtop: 20000003">
            <name xml:lang="en- GB">Ani mati on</name>
            <name xml:lang="fr">Dessin ani mé</name>
        </narrower>
    </subject>
    <headline>Loriot retrospective</headline>
    <description rol e="descrol e: dopesheet">
        Yesterday evening (November, 5) an exhibition opened in Berlin in
        honour of German humorist Vicco von Bülow, better known under the
        pseudonym "Loriot", to commemorate his 85th birthday. He was born
        November 12, 1923 in Brandenburg an der Havel and comes from an old
        German aristocratic family. He is most well-known for his cartoons,
        television sketches alongside late German actress Evelyn Hammann.
        <br />
    </description>
    <description rol e="descrol e: shotli st">
        Berlin, 21/12/2008
        <br />
        - vs. Vicco von Bülow entering exhibition
        <br />
        - vs. Loriot and media
        <br />
    </description>

```

```

- sot Vicco von Bülow
<br />
"Since 85 years I didn't succeed in pursuing a job that could be
called a profession."
<br />
- vs exhibition
<br />
- sot Irm Herrmann, actress
<br />
"Loriot is timeless. You always can watch him and I can always
laugh."
<br />
sot Ulrich Matthes, actor
<br />
"I would say: one of the great German classics. Goethe, Kleist,
Schiller, Thomas Mann, Loriot. That's the way I would say it."
<br />
</description>
</contentMeta>
<partMeta partid="Part1_ID" seq="1">
<icon href="http://www.example.com/video/Keyframes/407624.jpeg"/>
<timeDelim start="0" end="446" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:VoiceOver"/>
<description>Vicco von Bülow entering exhibition </description>
</partMeta>
<partMeta partid="Part2_ID" seq="2">
<icon href="http://www.example.com/video/Keyframes/407624-447.jpeg"/>
<timeDelim start="447" end="831" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:VoiceOver"/>
<description>Loriot and media </description>
</partMeta>
<partMeta partid="Part3_ID" seq="3">
<icon href="http://www.example.com/video/Keyframes/407624-832.jpeg"/>
<timeDelim start="832" end="1081" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:Interlocution"/>
<description>Vicco von Bülow interview</description>
</partMeta>
<partMeta partid="Part4_ID" seq="4">
<icon href="http://www.example.com/video/Keyframes/407624-1082.jpeg"/>
<timeDelim start="1082" end="1313" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:NaturalSound"/>
<description>Exhibition panorama </description>
</partMeta>
<partMeta partid="Part5_ID" seq="5">
<icon href="http://www.example.com/video/Keyframes/407624-1314.jpeg"/>
<timeDelim start="1314" end="1616" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:Interlocution"/>
<description>Irm Herrmann, actress, interview</description>
</partMeta>
<partMeta partid="Part6_ID" seq="6">
<icon href="http://www.example.com/video/Keyframes/407624-1617.jpeg"/>
<timeDelim start="1617" end="2109" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:VoiceOver"/>
<description>Ulrich Matthes, actor, in exhibition</description>
</partMeta>
<partMeta partid="Part7_ID" seq="7">
<icon href="http://www.example.com/video/Keyframes/407624-2110.jpeg"/>
<timeDelim start="2110" end="2732" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:Interlocution"/>
<description>Ulrich Matthes, actor, interview</description>
</partMeta>
<partMeta partid="Part8_ID" seq="8">
<icon href="http://www.example.com/video/Keyframes/407624-2733.jpeg"/>
<timeDelim start="2733" end="2774" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:VoiceOver"/>
<description>"I would say: one of the great German classics. Goethe, Kleist,
Schiller, Thomas Mann, Loriot. That's the way I would say it."</description>
</partMeta>
<contentSet>
<remoteContent href="http://www.example.com/video/407624.avi"
format="fmt:avi"
duration="111" durationunit="timeunit:seconds"
videocodec="vcdc:c155"
videoframerate="25"
videoaspectratio="16:9"/>
</contentSet>
</newsItem>

```



This page intentionally blank

11 Quick Start - Packages



Read the Quick Guide to G2 Basics
before this Quick Start.

11.1 Introduction

The ability to package together items of news content is important to news organisations and customers. Using packages, different facets of the coverage of a news story can be viewed in a named relationship, such as “Main Article”, “Sidebar”, “Background”. Another frequent application of packages is to aggregate content for news products, for example “Top Ten” news packages such as that illustrated below⁵.

UK



Figure 6: A Top Ten News Package displayed on the Web

Packages can range from simple collections on a common theme, to rich hierarchical structures.

NewsML-G2 is flexible in allowing a provider to package content that has already been published, or a package may be sent together with all of its content resources in a single News Message. See [18 Exchanging News: News Messages](#).

11.2 Packages and Links: the difference

The NewsML-G2 <link> property is a useful way to indicate optional supplementary resources that may be retrieved by the end-user when processing or consuming a G2 Item. Links should not be used as a lightweight method of packaging news; a G2 processor would not be able to distinguish between News

⁵ A description of how to create this type of package with ordered components can be found further on in this document.



Items with some optional resources, and News Items that are intended to be pseudo-packages using links. It is also a basic G2 rule that a News Item only conveys one piece of content.

By contrast, Packages:

- ❖ Express structure, allowing news to be packaged as a list, or as a named hierarchy of content resources.
- ❖ Have a mode property that enables the expression of a relationship between the components of a package group.

11.3 Document structure

The building blocks of the Package Item are the <packageItem> root element, with additional wrapping elements for metadata about the Package (itemMeta), metadata about the content (contentMeta) and the package content (groupSet). The top level (root) element <packageItem> attributes are:

```
<packageItem xml ns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:example.com,2008:UK-NEWS-TOPTEN:UK20081220098658" version="18">
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
```

This is followed by Catalog information:

```
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<catalogRef
  href="http://www.example.com/customer/cv/catalog4customers-1.xml" />
```

11.4 Item Metadata

The <itemMeta> wrapper contains properties that are aids to processing the package contents.

11.4.1 Profile

The <profile> element allows a provider to name a pre-arranged template or transformation stylesheet that can be used to process the package, for example “text and picture” could be the name of a template; “textpicture.xsl” would be an xsl stylesheet. The @versioninfo of a <profile> enables the template or stylesheet to be versioned:

```
<profile versioninfo="1.0.0.2">simple_text_with_picture.xsl</profile>
```

11.4.2 Item Metadata in full

```
<itemMeta>
  <itemClass qcode="ni nat: composite" />
  <provider qcode="nprov: REUTERS" />
  <versionCreated>2012-11-07T12:30:00Z</versionCreated>
  <firstCreated>2008-12-20T12:25:35Z</firstCreated>
  <pubStatus qcode="stat: usable" />
  <profile versioninfo="1.0.0.2">simple_text_with_picture.xsl</profile>
  <service qcode="svc: uktop">
    <name>Top UK News stories hourly</name>
  </service>
  <title>UK-TOPNEWS</title>
  <edNote>Updates the previous version</edNote>
  <signal qcode="sig: update" />
</itemMeta>
```

11.5 Content Metadata

The <contentMeta> wrapper in this example contains extended metadata about the person who compiled the package, including hours of duty and contact telephone number.

```
<contentMeta>
  <contributor jobtitle="staffjobs: cpe" qcode="mystaff: MDancer">
    <name>Maurice Dancer</name>
    <name>Chief Packaging Editor</name>
    <definition validto="2008-12-20T17:30:00Z">
      Duty Packaging Editor
    </definition>
    <note validto="2008-12-20T17:30:00Z">
```



```

    Available on +44 207 345 4567 until 17:30 GMT today
  </note>
</contributor>
<headline xml:lang="en">UK</headline>
</contentMeta>

```

11.6 Package Structure

A simple Package has a structure as shown in the example below. The top level for content of a Package Item is one and only one `<groupSet>` element, followed by at least one `<group>` structure containing one or more `<ItemRef>` references to content. The `<group>` structure may also be repeated, but this example has only one. The diagram below shows a skeleton of the XML elements in a simple package and a visualisation of the relationship that this structure creates:

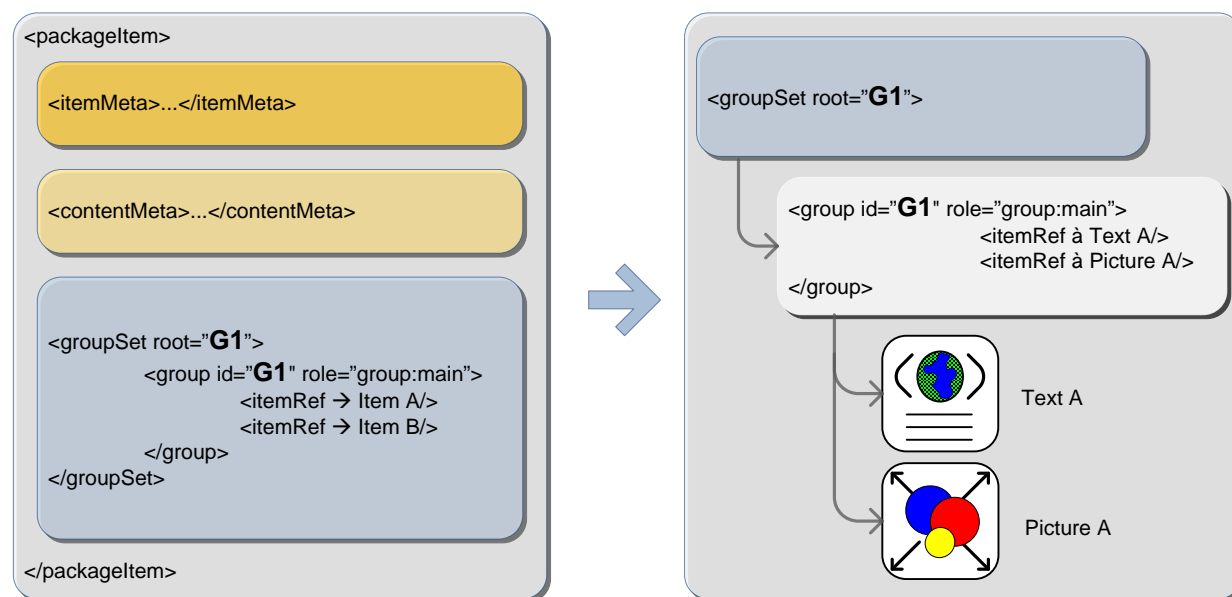


Figure 7: Top-level element view of a simple package, and (right) a visualisation of the structure

11.6.1 Group Set

The `<groupSet>` has a mandatory root attribute that is an idref to the primary child `<group>` element. The primary `<group>` element must identify itself using an `@id` that matches the `@root` of `<groupSet>`.

```
<groupSet root="G1">
```

11.6.2 Group

Although the `id` attribute is optional, in practice one must be provided to match the mandatory root attribute of the `<groupSet>`, even if there is only one `<group>`. If there is more than one `<group>` element, one and only one can be identified as the root group.

Group elements must also contain a `role` attribute to declare its role within the package structure. The role is a QCode, but a Scheme of Roles may typically contain values representing "main", "sidebar" or other editorial terms that express how the content is intended to be used in the package.

```
<group id="G1" role="group:main">
```

11.6.3 Item Reference

The `<itemRef>` element identifies an Item or a Web resource using `@href` and/or `@residref`. The IPTC recommends that Package Items should reference G2 Items if they are available (typically News Items) rather than other types of resource, such as "raw" news objects. Referring to other kinds of Web-accessible resource is allowed and is a legitimate use-case, however it has some disadvantages. Resources referred to in this way cannot be managed or versioned: if one of the resources is changed, the entire package may need to be re-compiled and sent, whereas a reference to a managed object such as a `<newsItem>` may refer to the latest (or a specific) version.



The example versions the referenced Items using @version, and gives processing or usage hints using @contenttype and @size. The @contenttype uses the registered IANA MIME type for a G2 News Item:

```
<itemRef resiHref="urn:newsml:iptc.org:20081007:tutorial-item-A"
  contentType="application/vnd.iptc.g2.newsitem+xml"
  size="2345">
```

The Item Reference includes properties from the referenced Item that have been extracted as an aid to processing:

```
<itemClass qcode="inat:text" />
<provider qcode="nprov:REUTERS"/>
<pubStatus qcode="stat:usable"/>
<title>Obama annonce son équipe</title>
<description role="drol:summary">Le rachat il y a deux ans de la
  propriété par Alan Gerry, magnat local de la télévision câblée, a
  permis l'investissement des 100 millions de dollars qui étaient
  nécessaires pour le musée et ses annexes, et vise à favoriser le
  développement touristique d'une région frappée par le chômage.
</description>
</itemRef>
```

11.7 Complete Listing of a Simple Package

The Listing below is given for completeness.

LISTING 6 Simple NewsML-G2 Package

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<packageItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:example.com,2008:UK-NEWS-TOPTEN:UK20081220098658" version="18">
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef
    href="http://www.example.com/customer/cv/catalog4customers-1.xml" />
  <itemMeta>
    <itemClass qcode="inat:composite" />
    <provider qcode="nprov:REUTERS" />
    <versionCreated>2012-11-07T12:30:00Z</versionCreated>
    <firstCreated>2008-12-20T12:25:35Z</firstCreated>
    <pubStatus qcode="stat:usable" />
    <profile versionInfo="1.0.0.2">simple_text_with_picture.xsl</profile>
    <service qcode="svc:uktop">
      <name>Top UK News stories hourly</name>
    </service>
    <title>UK-TOPNEWS</title>
    <edNote>Updates the previous version</edNote>
    <signal qcode="sig:update" />
  </itemMeta>
  <contentMeta>
    <contributor jobTitle="staffjobs:cpe" qcode="mystaff:MDancer">
      <name>Maurice Dancer</name>
      <name>Chief Packaging Editor</name>
      <definition validto="2008-12-20T17:30:00Z">
        Duty Packaging Editor
      </definition>
      <note validto="2008-12-20T17:30:00Z">
        Available on +44 207 345 4567 until 17:30 GMT today
      </note>
    </contributor>
    <headline xml:lang="en">UK</headline>
  </contentMeta>
  <groupSet root="G1">
    <group id="G1" role="group:main">
      <itemRef resiHref="urn:newsml:iptc.org:20081007:tutorial-item-A"
        contentType="application/vnd.iptc.g2.newsitem+xml"
        size="2345">
        <itemClass qcode="inat:text" />
        <provider qcode="nprov:REUTERS"/>
        <pubStatus qcode="stat:usable"/>
        <title>Obama annonce son équipe</title>
        <description role="drol:summary">Le rachat il y a deux ans de la
          propriété par Alan Gerry, magnat local de la télévision câblée, a
          permis l'investissement des 100 millions de dollars qui étaient
```



```

        nécessaires pour le musée et ses annexes, et vise à favoriser le
        développement touristique d'une région frappée par le chômage.
    </description>
</itemRef>
<itemRef resi dref="urn: newsml : iptc. org: 20081007: tutorial -i tem- B"
    content type="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="300039">
    <itemClass qcode="ni nat: pi ctur e" />
    <provider qcode="nprov: REUTERS"/>
    <pubStatus qcode="stat: usable"/>
    <title>Barack Obama arrive à Washington</title>
    <description role="drol: capti on">Si nous avons aujourd' hui un
        afro- améri cain et une femme dans la course à la présidence.
    </description>
</itemRef>
</group>
</groupSet>
</packageItem>

```

11.8 Hierarchical Package Structure

Hierarchies of Groups and Item References can be created by adding multiple Groups to Packages and using <groupRef>, to reference other Groups by @idref, as illustrated by the following diagram:

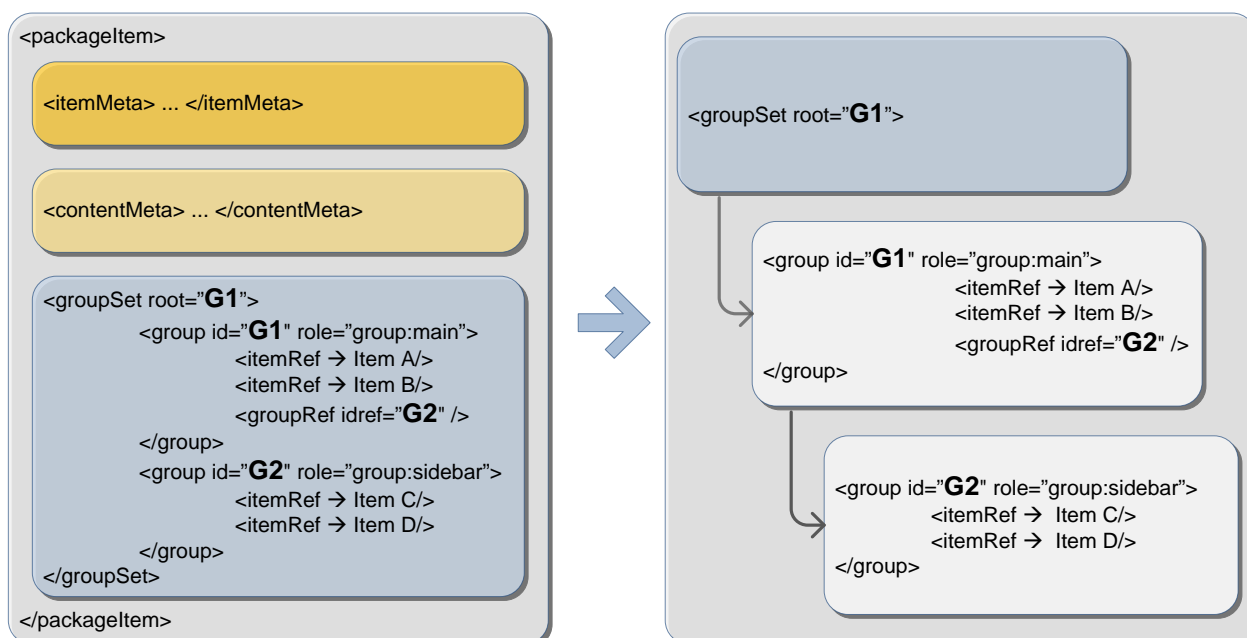


Figure 8: Code outline of hierarchical package with two groups, visualising parent-child structure (right)

The code listing below shows how such a hierarchical package would be fully expressed in XML in a NewsML-G2 Group Set:

LISTING 7 Group Set example showing Hierarchical Package Structure

```

<groupSet root="G1">
  <group id="G1" role="group: main">
    <itemRef resi dref="urn: newsml : iptc. org: 20081007: tutorial -i tem- A"
      content type="appli cation/vnd. iptc. g2. newsi tem+xml "
      size="2345">
      <itemClass qcode="ni nat: text" />
      <title>Obama annonce son équipe</title>
    </itemRef>
    <itemRef resi dref="urn: newsml : iptc. org: 20081007: tutorial -i tem- B"
      content type="appli cation/vnd. iptc. g2. newsi tem+xml "
      size="300039">
      <itemClass qcode="ni nat: pi ctur e" />
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
    <groupRef idref="G2" />
  </group>
  <group id="G2" role="group: si debar">

```

```
<itemRef resioref="urn:newsml:iptc.org:20081007:tutorial-item-C"
  contentType="application/vnd.iptc.g2.newsitem+xml"
  size="1503">
  <itemClass qcode="ni nat: text" />
  <title>Clinton reprend son rôle de chef de la santé</title>
</itemRef>
<itemRef resioref="urn:newsml:iptc.org:20081007:tutorial-item-D"
  contentType="application/vnd.iptc.g2.newsitem+xml"
  size="350280">
  <itemClass qcode="ni nat: picture" />
  <title>Hillary Clinton à une rassemblement à New York</title>
</itemRef>
</group>
</groupSet>
```

In the example, the “root” group is identified as the group with id=“G1”. This group has a role of “main” and consists of a text story and a picture of Barack Obama. The group with id=“G2” has the role of “sidebar” and contains a text and picture of Hillary Clinton. It is referenced by a <groupRef> in Group G1.

11.9 List Type Package Structure

The @mode indicates the relationship between components of a group using one of three values from the IPTC Package Group Mode NewsCodes (recommended Scheme Alias “pgrmod”):

- ❖ pgrmod:bag – an unordered collection of components, for example different components of a web news page with no special order, as in the example below. This is the default @mode.
- ❖ pgrmod:seq – denotes a sequential package group set in descending order, for example a “Top Ten” list: each sub-group would provide references to a text article and a related picture.
- ❖ pgrmod:alt – an unordered collection. Each sub-group is an alternative to its peer groups in the set, for example coverage of a news event supplied in different languages.

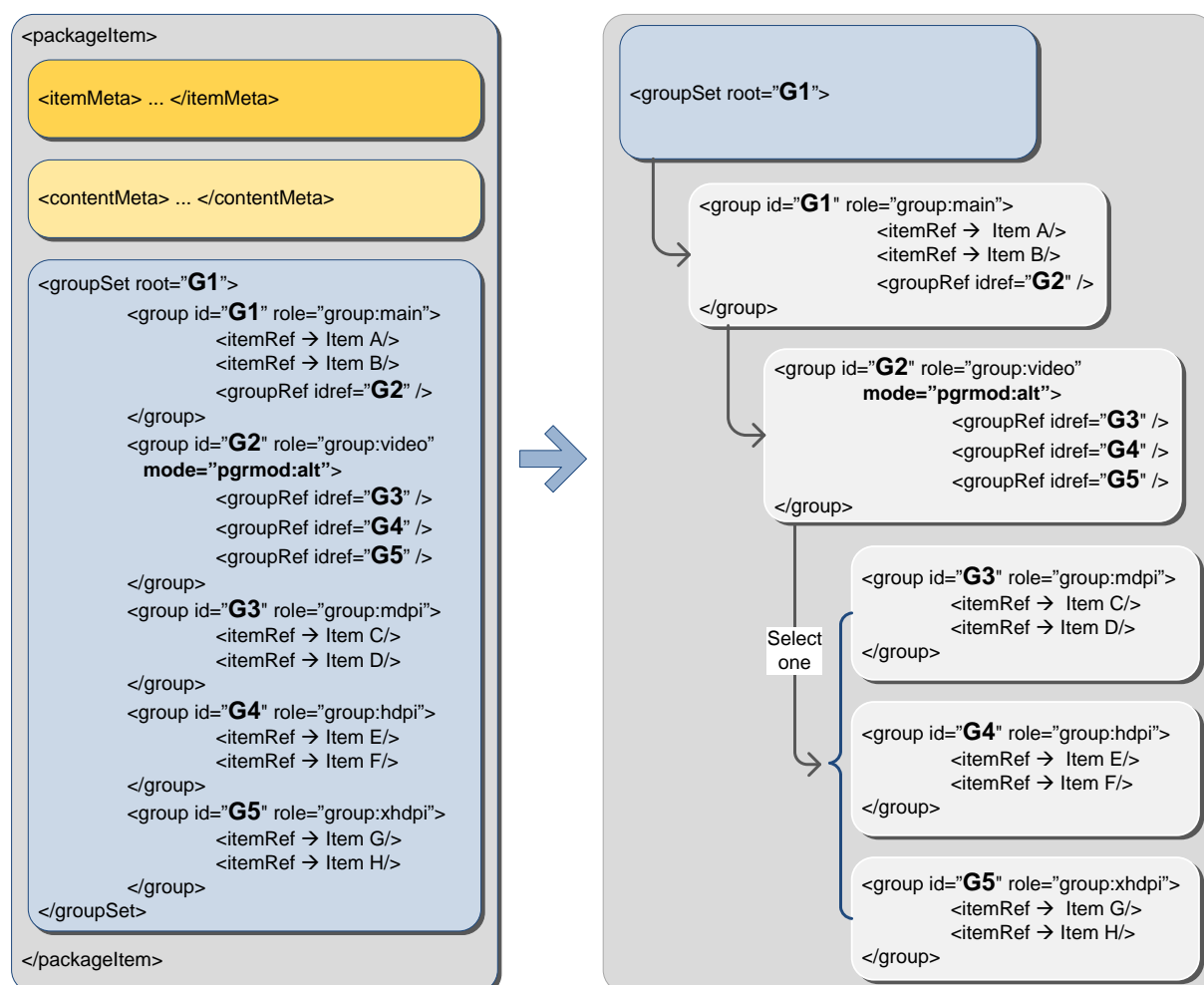


Figure 9: Code skeleton view of package with alternative components, with visualisation of structure



The diagram above shows a package containing two Items in the root group, and a group reference to a “group of groups” with package mode set to “alt” indicating that the child groups contain alternative content. The example uses groups of associated video suitable for different Android device screen resolutions as indicated by the @role of each group.

The code overview shows the root group referencing the two Items and the <groupRef> element referencing the group with @id “G2”. Group G2 has its package mode set to “alt” and its components are references to alternate groups G3, G4 and G5, which reference videos at the required resolution for each screen type.

The right-hand image in the diagram is a visual representation of the relationship expressed through this package structure.

Note the <group> that has its Mode set to “alt” – not the “main” group but the second group with @id “G2”. The components of this group are alternatives: each references a group containing the video content. The code example below shows how this relationship is fully expressed in NewsML-G2:

LISTING 8 Group Set example showing an “alt” Package Mode

```
<groupSet root="G1">
  <group id="G1" role="group: main">
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-A"
      contenttype="application/vnd. iptc. g2. newsitem+xml"
      size="2345">
      <itemClass qcode="ni nat: text" />
      <title>Obama annonce son équipe</title>
    </itemRef>
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-B"
      contenttype="application/vnd. iptc. g2. newsitem+xml"
      size="1503">
      <itemClass qcode="ni nat: text" />
      <title>Clinton reprend son rôle de chef de la santé</title>
    </itemRef>
    <groupRef idref="G2" />
  </group>
  <group id="G2" role="group: video" mode="pgrmod: alt">
    <groupRef idref="G3" />
    <groupRef idref="G4" />
    <groupRef idref="G5" />
  </group>
  <group id="G3" role="group: mdpi">
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-C"
      contenttype="video/mp4" width="480" height="320">
      <itemClass qcode="ni nat: video" />
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-D"
      contenttype="video/mp4" width="480" height="320">
      <itemClass qcode="ni nat: video" />
      <title>Hillary Clinton à une rassemblement à New York</title>
    </itemRef>
  </group>
  <group id="G4" role="group: hdpi">
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-E"
      contenttype="video/mp4" width="720" height="480">
      <itemClass qcode="ni nat: video" />
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-F"
      contenttype="video/mp4" width="720" height="480">
      <itemClass qcode="ni nat: video" />
      <title>Hillary Clinton à une rassemblement à New York</title>
    </itemRef>
  </group>
  <group id="G5" role="group: xhdpi">
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-G"
      contenttype="video/mp4" width="960" height="640">
      <itemClass qcode="ni nat: video" />
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial-item-H"
      contenttype="video/mp4" width="960" height="640">
      <itemClass qcode="ni nat: video" />
      <title>Hillary Clinton à une rassemblement à New York</title>
    </itemRef>
  </group>
</groupSet>
```

```
</group>
</groupSet>
```

11.10 A Sequential “Top Ten” Package

The screenshot at the start of this Guide shows a “Top Ten” list of news items in order of importance. The package mode of “seq” indicates that the components are in descending order and a code skeleton and visual representation of the package structure is shown in the diagram below:

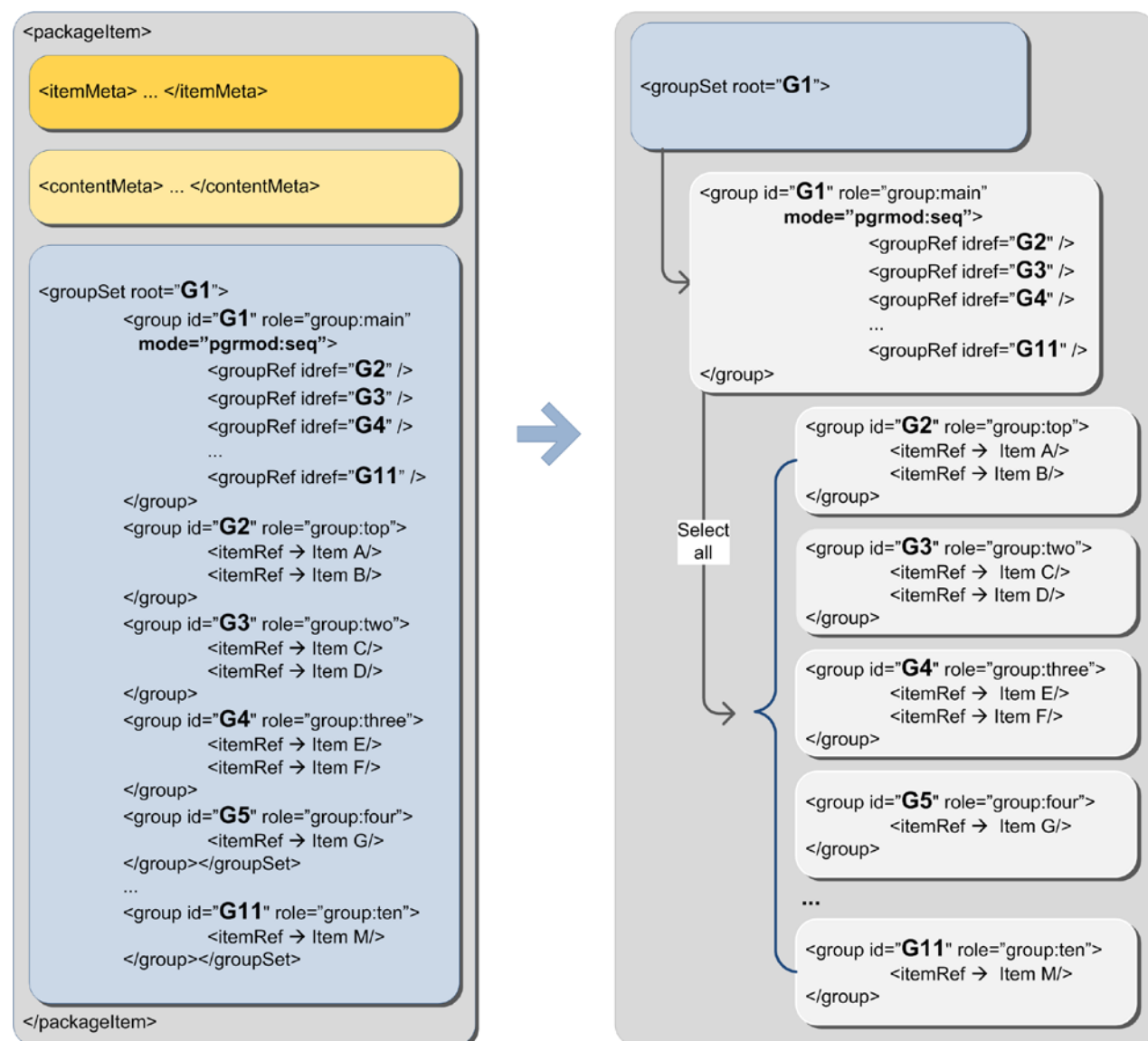


Figure 10: Code skeleton of a sequential mode package and (right) the resulting relationship structure

Note how the `<group>` sets the Mode for its components, in this case the component group references of the “main” group are sequentially ordered. The relationship is fully-expressed in XML in NewsML-G2 as shown below:

LISTING 9 Group Set example showing a “seq” Package Mode

```
<groupSet root="G1">
  <group id="G1" role="group:main" mode="pgrmod:seq">
    <groupRef idref="G2" />
    <groupRef idref="G3" />
    <groupRef idref="G4" />
    <groupRef idref="G5" />
    <groupRef idref="G6" />
    <groupRef idref="G7" />
    <groupRef idref="G8" />
    <groupRef idref="G9" />
```



```

<groupRef idref="G10" />
<groupRef idref="G11" />
</group>
<group id="G2" role="group: top" >
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- A"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="3452">
    <itemCl ass qcode="ni nat: text" />
    <provider qcode="nprov: REUTERS"/>
    <pubStatus qcode="stat: usable"/>
    <title>Bank cuts interest rates to record low</title>
    <description role="drol: summary">London (Reuters) - The Bank of England cut
      interest rates by half a percentage point on Thursday to a record low of
      1.5 percent and economists expect it to cut again in February as it
      battles to prevent Britain from falling into a deep slump.
    </description>
  </itemRef>
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- B"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="230003">
    <itemCl ass qcode="ni nat: pi ctur e" />
    <provider qcode="nprov: REUTERS" />
    <pubStatus qcode="stat: usable"/>
    <title>BoE Rate Deci sion</title>
  </itemRef>
</group>
<group id="G3" role="group: two" >
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- C"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="2345">
    <itemCl ass qcode="ni nat: text" />
    <provider qcode="nprov: REUTERS"/>
    <pubStatus qcode="stat: usable"/>
    <title>Government denies it will print more cash</title>
    <description role="drol: summary">London (Reuters) - Chancellor
      Alistair Darling dismissed reports on Thursday that the government was about
      to boost the money supply to ease the impact of recession.
    </description>
  </itemRef>
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- D"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="24065">
    <itemCl ass qcode="ni nat: pi ctur e" />
    <provider qcode="nprov: REUTERS" />
    <pubStatus qcode="stat: usable"/>
    <title>Sterling notes and coin</title>
  </itemRef>
</group>
<group id="G4" role="group: three" >
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- E"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="2345">
    <itemCl ass qcode="ni nat: text" />
    <provider qcode="nprov: REUTERS"/>
    <pubStatus qcode="stat: usable"/>
    <title>Rugby's Mike Tindall banned for drink-driving</title>
    <description role="drol: summary">London (Reuters) - England rugby player Mike
      Tindall was banned from driving for three years and fined £500 on Thursday
      for his second drink-drive offence.
    </description>
  </itemRef>
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- F"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="25346">
    <itemCl ass qcode="ni nat: pi ctur e" />
    <provider qcode="nprov: REUTERS" />
    <pubStatus qcode="stat: usable"/>
    <title>Mike Tindall in rugby action for Engl and</title>
  </itemRef>
</group>
<group id="G5" role="group: four" >
  <itemRef residref="urn: newsml: iptc. org: 20081007: tutorial- item- G"
    contenttype="appli cation/vnd. iptc. g2. newsi tem+xml "
    size="3654">
    <itemCl ass qcode="ni nat: text" />
    <title>Crunch forces employees to work unpaid overtime</title>
  </itemRef>
</group>
<group id="G6" role="group: fi ve" >

```




```

<itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-H"
  contentType="application/vnd.iptc.g2.newsitem+xml"
  size="5123">
  <itemClass qcode="ni nat: text" />
  <title>Government warns of tax fraudsters</title>
</itemRef>
</group>
<group id="G7" role="group: six" >
  <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-I"
    contentType="application/vnd.iptc.g2.newsitem+xml"
    size="4323">
    <itemClass qcode="ni nat: text" />
    <title>Nissan to cut 1,200 jobs at Sunderland plant</title>
  </itemRef>
</group>
<group id="G8" role="group: seven" >
  <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-J"
    contentType="application/vnd.iptc.g2.newsitem+xml"
    size="3122">
    <itemClass qcode="ni nat: text" />
    <title>Sainsbury sales tops forecast</title>
  </itemRef>
</group>
<group id="G9" role="group: eight" >
  <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-K"
    contentType="video/mp4-480x320"
    size="322443">
    <itemClass qcode="ni nat: video" />
    <title>Cause of wind turbine damage unknown</title>
  </itemRef>
</group>
<group id="G10" role="group: nine" >
  <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-L"
    contentType="application/vnd.iptc.g2.newsitem+xml"
    size="4123">
    <itemClass qcode="ni nat: text" />
    <title>Muslims warn Gaza crisis could provoke extremism</title>
  </itemRef>
</group>
<group id="G11" role="group: ten" >
  <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-M"
    contentType="application/vnd.iptc.g2.newsitem+xml"
    size="8192">
    <itemClass qcode="ni nat: text" />
    <title>Banks hiring young Britons to prepare for upturn</title>
  </itemRef>
</group>
</groupSet>

```

11.11 Package Processing Considerations

11.11.1 Other NewsML-G2 Items

In the above examples, the referenced resources in the package have been News Items, but <itemRef> may also refer to other Items, such as Package Items. The following example of <itemRef> shows how a Package Item can be used as part of a Package Item. This type of “Super Package” could be used to send a “Top Ten” package (a themed list of news) where each referenced item is also a package consisting of references to the text, picture and video coverage of each news story.

The advantage of using this “package of packages” approach is that it promotes more efficient re-use of content. Once created, any of the “sub-packages” can be easily referenced by more than one “super-package”: a package about a given story could be used by both “Top News This Hour” and by “Today’s Top News”. If the individual News Items that make up a sub-package were to be referenced directly, these references have to be assembled each time the story is used, either by software or a journalist, which would be less efficient.

As these sub-packages are managed objects, we use @residref to identify and locate the referenced items. Each referenced item may be a Package Item, shown by the Item Class of “composite” and the Content Type of “application/vnd.iptc.g2.packageitem+xml”. Each <itemRef> would then resemble the following:

```
<itemRef residref="tag:afp.com,2008:TX-PAR:20080529:JYC99"
  contentType="application/vnd.iptc.g2.packageitem+xml"
  size="28047">
  <itemClass qcode="inat:composite" />
  <provider qcode="nprov:AFP"/>
  <pubStatus qcode="stat:usable"/>
  <title>Tiger Woods cherche son retour</title>
  <description role="drol:summary">Tiger Woods lorem ipsum dolor sit amet,
    consectetur adipiscing elit. Etiam feugiat. Pellentesque ut enim eget
    eros volutpat consectetur. Quisque sollicitudin, tortor ut dapibus
    porttitor, augue velit vulputate eros, in tempus orci nunc vitae nunc.
    Nam et lacus ut leo convallis posuere. Nullam risus.
  </description>
</itemRef>
```

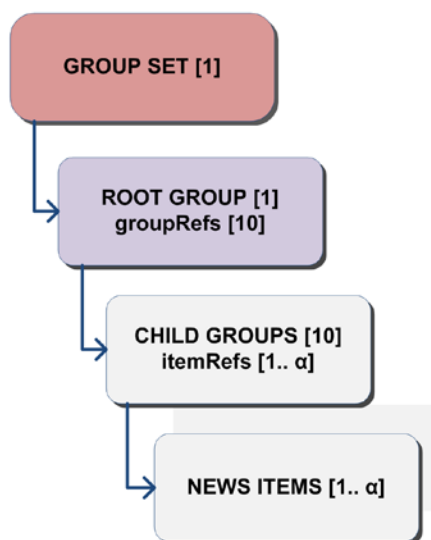
11.11.2 Facilitating the Exchange of Packages

There needs to be some consideration of how such a “Super Package” should be processed by the receiver. The power and flexibility inherent in NewsML-G2 Packages could lead to confusion and processing complexity unless provider and receiver agree on a method for specifying the structure of packages and signalling this to the receiving application. Processing hints such as the <profile> property (described above) intended to help resolve this issue.

In the example below, we maintain flexibility and inter-operability with potential partner organisations by defining any number of standard package “templates” – termed Profiles – for the Package, among other processing hints. Partners would agree in advance on the Profiles and rules for processing them. All that the provider then needs to do is place the pre-arranged Profile name, or the name of a transformation script, in the <profile> property.

Package profiles could be represented as diagrams like those shown below:

“TOP TEN” LIST PROFILE



MULTIMEDIA STORY PROFILE

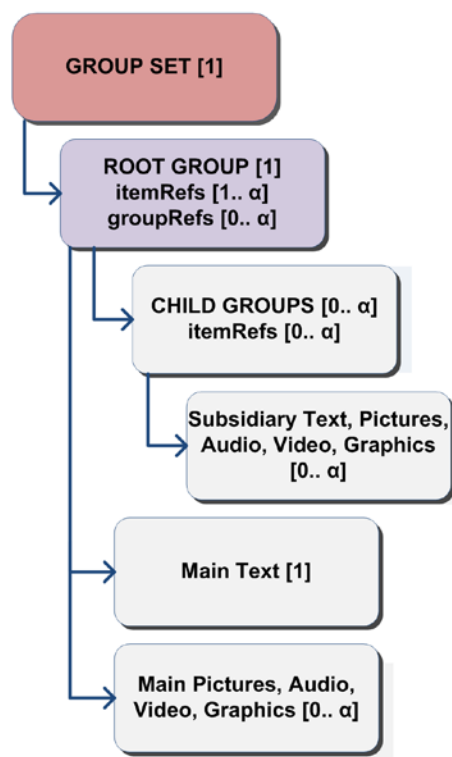


Figure 11: Diagrams of Package Profiles. The numbers in brackets indicate the required items

In this example, the Profile Name is intended to be a signal to the processor that references to each member of the Top Ten list are placed in their own group, and that we create our Top Ten list in the “root” group of the Package Item as an ordered list of <groupRef> elements. (as in the “Top Ten” list profile shown in **Figure 11**)

The properties in <itemMeta> that can be used to provide information on processing are:



<generator>, a versioned string denoting the name of the process or service that created the package:

```
<generator versionInfo="3.0">MyNews Top Ten Packager</generator>
```

<profile>, as discussed, sets the template or transformation stylesheet of the package

```
<profile versionInfo="1.0.0.2">ranked_idref_list</profile>
```

<signal> is a QCode type property that instructs the receiver to perform any required actions upon receiving the Item. An <edNote> may contain natural-language instructions, if necessary, and a <link> property denotes the previous version of the package.

```
<signal qcode="action: replacePrev" />
<edNote>Replace the previous package</edNote>
<link
  rel="i rel: previousVersion"
  resIdref="tag: example. com, 2008: UK- NEWS- TOPTEN: UK20081220098658"
  version="1"
/>
```

12 Concepts and Concept Items

12.1 Introduction

Concepts in NewsML-G2 are a method of describing real-world entities, such as people, events and organisations, and also to describe thoughts or ideas: abstract notions such as subject classifications, facial expressions. Using concepts, we can classify news, and the entities and ideas found in news, to make the content more accessible and relevant to people's particular information needs.

Content originators who make up the IPTC membership constantly strive to increase the value proposition of their products. The need to extract and properly express the meaning of news using concepts is a major reason for moving to NewsML-G2.

Clear and unambiguously-defined concepts enable receivers of information to categorize and otherwise handle news more effectively, routing content and archiving it accurately and quickly using automated processes.

NewsML-G2 Concepts are powerful because they bring meaning to news content in a way that can be understood by humans and processed by machines. The concept model aligns with work being done at the W3C and elsewhere to realize the Semantic Web.

Concepts are conveyed individually in Concept Items, or (more commonly) are collected as groups of Concepts in Knowledge Items. These can be collections with a common purpose, such as Controlled Vocabularies.

This Chapter gives details of the Concept element that is common to both types of Item, and also describes the Concept Item. The Chapter [13 Knowledge Items](#), succeeding this one, described Knowledge Items in detail.

Concepts are also used to convey event information, which is described in detail in [EventsML-G2](#).

12.2 What is a Concept?

A NewsML-G2 Concept is anything about which we can express knowledge in some formal way, and which may also have a named relationship with other concepts:

- ❖ “Mario Draghi” is a concept about which, or whom, we can express knowledge, for example, date of birth (September 3, 1947), job title (President of the European Central Bank).
- ❖ “The European Central Bank” is a concept. It has an address, a telephone number, and other inherent characteristics of an organisation.
- ❖ We can express a named relationship: “Mario Draghi” *is a member of* “The European Central Bank” NewsML-G2 concept expressions thus conform with an RDF triple of subject, predicate and object

Concepts are either global in scope, when they are identified by a URI using a @uri, optionally taking the format of a QCode attribute @qcode, or their scope is local to the containing document when identified by a string value using @literal (where permitted) The use of @literal identifier is a special case that matches the identifier of an <assert> in the NewsML-G2 document that contains a localised concept structure. See 20.2 The Assert Wrapper for more details. This Chapter describes Concepts identified by QCodes and URIs.

12.3 Creating Concepts – the <concept> element

The <concept> element contains the properties that express the concept in detail and identify it so that it can be used and re-used:

12.3.1 Concept ID <conceptId>

A concept MUST contain a <conceptId> which takes the form of a QCode (@qcode) attribute. Optionally the full URI may be added using a @uri. If URI resolved from @qcode is not the same as the @uri value, then the URI resolved from the @qcode takes precedence. Optionally, this can be refined using date-time for @created and @retired.

```
<concept>
  <conceptId created="2009-01-01T12:00:00Z" qcode="foo:bar" />
  ...
</concept>
```

When a concept is retired by use of the @retired attribute of <conceptId>, the authority behind the concept is indicating that it is no longer actively using this concept (for example it may have been merged with another concept), but resources that were created before the change must continue to be able to resolve the concept.

12.3.2 Concept Name <name>

A concept MUST contain at least one <name>, a natural language name for the concept, with optional attributes of @xml:lang and @dir (text direction):

```
<concept>
  <conceptId qcode="foo:bar" />
  <name>Mario Draghi</name>
</concept>
```

Concepts are designed to be useable in multiple languages:

```
<concept>
  <conceptId created="2000-10-30T12:00:00+00:00" qcode="medtop:01000000" />
  <type qcode="cpnat:abstract" />
  <name xml:lang="en-GB">arts, culture and entertainment</name>
  <name xml:lang="de">Kultur, Kunst, Unterhaltung</name>
  <name xml:lang="fr">Arts, culture, et spectacles</name>
  <name xml:lang="es">arte, cultura y espectáculos</name>
  <name xml:lang="ja-JP">文化</name>
  <name xml:lang="it">Arte, cultura, intrattenimento</name>
</concept>
```

12.3.3 Concept Type

The optional <type> element expresses the “nature of the Concept”, for example, using the recommended IPTC Concept Nature NewsCodes to identify this concept is of type “abstract”. We can also use <related> to extend this notion into further characteristics of the concept (see [Relationships between Concepts](#), below).

<type> demonstrates the use of the subject, predicate, object triple of RDF to express a named relationship with another concept; <type> can only express one kind of relationship – “is a(n)”. It is used to express the most obvious, or primary, inherent characteristic of a concept, as in:

“arts, culture and entertainment” (Subject) “is an” (Predicate) “abstract concept” (Object):

```
<type qcode="cpnat:abstract" />
```

The current types agreed by the IPTC and contained in the “concept nature” CV at <http://cv.iptc.org/newsCodes/cpnature/> are:

- ❖ abstract concept (cpnat:abstract),
- ❖ person (cpnat:person),
- ❖ organisation (cpnat:organisation),
- ❖ geopolitical area (cpnat:geoArea),
- ❖ point of interest (cpnat:poi),
- ❖ object (cpnat:object)
- ❖ event (cpnat:event).

12.3.4 Concept Definition

The optional <definition> element allows more extensive natural language information with some mark-up, if required. Block type elements may use an optional @role QCode to differentiate repeating Definition statements such as “summary” and “long”:

```
<definition xml:lang="en-GB" role="drol:short">
  Matters pertaining to the advancement and refinement of the human mind,
  of interests, skills, tastes and emotions
</definition>
```

Note that although much of this information could be, and may be, duplicated in machine-readable XML, it is still useful to carry some core information in human-readable form.

12.3.5 Note

The <note> element may be used to add supplemental natural-language information on the concept as a block of text with some optional mark-up, again with an optional @role:

```
<note role="nrol:supp">
  This is a top-level concept from the IPTC Media Topic NewsCodes
</note>
```

12.4 Conveying Concepts: the Concept Item structure

A Concept Item conveys knowledge about a single concept, whether a real-world entity such as a person, or an abstract concept such as a subject. It shares the basic structure of all NewsML-G2 Items and therefore uses the same methods for identification, versioning and conformance levels.

Item Metadata is mandatory and contains the mandatory properties for Item Class, Provider and Version Created (note that Publication Status is optional but the Item's publication status must be assumed to be the default "usable" if the property is absent).

Content Metadata is optional and is not included in this example:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20080229:ncdci-subjectcode"
  version="101018123521"
  standard="NewsML-G2"
  standardversion="2.18"
  catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<rightsInfo>
  <copyrightHolder>
    <name>IPTC - International Press Telecommunications Council, 20 Garrick Street,
London WC2E 9BT, UK</name>
  </copyrightHolder>
  <copyrightNotice>Copyright 2008, IPTC, www.iptc.org, All
Rights Reserved</copyrightNotice>
</rightsInfo>
<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2012-11-07T12:35:21+01:00</versionCreated>
  <firstCreated>2008-02-29T12:00:00+00:00</firstCreated>
  <pubStatus qcode="stat:usable" />
  <title xml:lang="en">Concept Item delivering a
concept requested from the IPTC Media Topic NewsCodes</title>
</itemMeta>
```

Note the <itemClass> property for a Concept Item must use the IPTC Concept Item Nature NewsCodes with a recommended Scheme Alias of "cinat" and denotes this Item conveys a NewsML-G2 Concept.

12.4.1 Completed Concept Item

This example is a Concept Item that describes one of the IPTC Media Topic NewsCodes:

LISTING 10 Abstract Concept conveyed in a NewsML-G2 Concept Item

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20080229:ncdci-subjectcode"
  version="101018123521"
  standard="NewsML-G2"
  standardversion="2.18">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <rightsInfo>
    <copyrightHolder>
      <name>IPTC - International Press Telecommunications Council, 20
Garrick Street, London WC2E 9BT, UK</name>
    </copyrightHolder>
    <copyrightNotice>Copyright 2008, IPTC, www.iptc.org, All
Rights Reserved</copyrightNotice>
  </rightsInfo>
</conceptItem>
```



```

</rightsInfo>
<itemMeta>
  <itemClass qcode="ci nat: concept" />
  <provider qcode="nprov: IPTC" />
  <versionCreated>2012-11-07T12:35:21+01:00</versionCreated>
  <firstCreated>2008-02-29T12:00:00+00:00</firstCreated>
  <pubStatus qcode="stat: usable" />
  <title xml:lang="en">Concept Item delivering a
    concept requested from the IPTC Media Topic NewsCodes</title>
</itemMeta>
<concept>
  <conceptId created="2000-10-30T12:00:00+00:00" qcode="medtop: 01000000" />
  <type qcode="cpnat: abstract" />
  <name xml:lang="en-GB">arts, culture and entertainment</name>
  <name xml:lang="de">Kultur, Kunst, Unterhaltung</name>
  <name xml:lang="fr">Arts, culture, et spectacles</name>
  <name xml:lang="es">arte, cultura y espectáculos</name>
  <name xml:lang="ja-JP">文化</name>
  <name xml:lang="it">Arte, cultura, intrattenimento</name>
  <definition xml:lang="en-GB">Matters pertaining to the advancement and refinement
    of the human mind, of interests, skills, tastes and emotions</definition>
  <definition xml:lang="de">Sachverhalte, die die Veränderung und Weiterentwicklung
    des menschlichen Geistes, der Interessen, des Geschmacks, der Fähigkeiten und
    der Gefühle betreffen.</definition>
  <definition xml:lang="fr">Tout ce qui est relatif à la création d'œuvres, au
    développement des facultés intellectuelles, et à leur représentation
    publique</definition>
  <definition xml:lang="es">Asuntos pertinentes al avance y refinamiento de la mente
    humana, intereses, habilidades, gustos y emociones.</definition>
  <definition xml:lang="ja-JP">
    人間の精神や興味、技能、嗜好、感情の進歩や洗練に係る事柄</definition>
  <definition xml:lang="it">Creazione e rappresentazione dell'opera d'arte, gli
    Interessi intellettuali, il gusto e le emozioni umane</definition>
  <note xml:lang="en-GB" role="nrol: supp">
    This is a top-level concept from the IPTC Subject NewsCodes
  </note>
</concept>
</conceptItem>

```

12.5 Concepts for real-world entities

For each of the types of named entities agreed by the IPTC: person, organisation, geographical area, point of interest, object and event, there is a specific group of additional properties. The following example is a Concept Item for a person.

12.5.1 Document Structure

The document structure is as previously described, with a root <conceptItem> element and <itemMeta>. The <contentMeta> element is optional and may only contain Administrative metadata properties, such as <contentModified> (not included in the example)

```

<?xml version="1.0" encoding="UTF-8"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20080229:ncdci-person"
  version="1010181123618"
  standard="NewsML-G2"
  standardversion="2.18"
  catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<rightsInfo>
  <copyrightHolder>
    <name>IPTC - International Press Telecommunications Council, 20
      Garrick Street, London WC2E 9BT, UK</name>
  </copyrightHolder>
  <copyrightNotice>Copyright 2008, IPTC, www.iptc.org, All
    Rights Reserved</copyrightNotice>
</rightsInfo>
<itemMeta>
  <itemClass qcode="ci nat: concept" />
  <provider qcode="nprov: IPTC" />
  <versionCreated>2012-11-07T12:38:18Z</versionCreated>
  <firstCreated>2008-12-29T11:00:00Z</firstCreated>
  <pubStatus qcode="stat: usable" />

```

```
<title xml:lang="en">Concept Item describing Mario Draghi</title>
</itemMeta>
```

12.5.2 Top-level concept details

The <concept> wrapper starts with the properties common to all types of concepts:

```
<concept>
  <conceptId created="2009-01-10T12:00:00Z" qcode="people:329465" />
  <type qcode="cpnat:person" />
  <name xml:lang="en-GB">Mario Draghi</name>
  <definition xml:lang="en-GB" role="drol:biog">Mario Draghi, born 3 September 1947,
    is an Italian banker and economist who succeeded Jean-Claude Trichet as the
    President of the European Central Bank on 1 November 2011. He was previously
    the governor of the Bank of Italy from January 2006 until October 2011. In 2014
    Forbes nominated Draghi 9th most powerful person in the world.<br />
  </definition>
  <note xml:lang="en-GB" role="nrol:disambiguation">
    Not Mario D'roggia, international powerboat racer
  </note>
  <related rel="relation:occupation" qcode="jobtypes:puboff" />
  <sameAs type="cpnat:person" qcode="pers:567223">
    <name>DRAGHI, Mario</name>
  </sameAs>
  ...
</concept>
```

Note the inclusion of Concept Relationship properties: the <related> element indicates that the person who is the subject of the concept “has occupation of” the related concept expressed in by the @qcode “jobtypes:puboff”. The <sameAs> element indicates that this concept is the same as AFP’s (note: fictitious) concept expressed by the @qcode “pers:567223”.

12.5.3 Person details

The <personDetails> element is a container for additional properties that are specifically designed to convey information about people:

12.5.3.1 Born <born> and Died <died>

The date of birth and date of death of the person, for example:

```
<born>1947-09-03</born>
```

The data type is “TruncatedDateTime”, which means that the value is a date, with an optional time part. The date value may be truncated from the right to a minimum of YYYY. If used, the time must be present in full, with time zone, and ONLY in the presence of the full date.

```
<born>1947</born>
```

12.5.3.2 Affiliation <affiliation>

An affiliation of the person to an organisation.

```
<affiliation type="orgnat:employer" qcode="org:ECB">
  <name>European Central Bank</name>
</affiliation>
```

Note that the @type refers to the type of organisation – not the type of relationship with the person. In the example we use scheme “orgnat” to describe the Nature of the Organisation as a Bank.

12.5.3.3 Contact Info <contactInfo>

Contact information associated with the person. The <contactInfo> element wraps a structure with the properties outlined below. A “person” concept may have many instances of <contactInfo>, each with @role indicating their purpose, or example work or home. These are controlled values, so a provider should create their own CV of address types if required.

Each of the child elements of <contactInfo> may be repeated as often as needed to express different @roles, for example different “work” and “personal” email addresses etc.

Property Name	Element	Type	Notes/Example
Email Address	<email>	Electronic Address	An "Electronic Address" type allows the expression of @role (QCode) to qualify the information, for example <pre><email role="addressrole: office"> info@ecb. eu </email></pre>
Instant Message Address	<im>	Electronic Address	<pre><im role="imsrvc: reuters"> j c. trichet. ecb. eu@reuters. net </im></pre>
Phone Number	<phone>	Electronic Address	
Fax Number	<fax>	Electronic Address	
Web site	<web>	IRI	<pre><web role="webrole: corporate"> www. ecb. eu </web></pre>
Postal Address	<address>	Address	See below for details of Address properties. The Address may have a @role to denote the type of address is contains (e.g. work, home) and may be repeated as required to express each address @role.
Other information	<note>	Block	Any other contact-related information, such as "annual vacation during August"

12.5.3.4 Postal address <address>

The Address Type property may have a @role to indicate its purpose, The following table shows the available child properties. Apart from <line>, which is repeatable, each element may be used once for each <address>

Property Name	Element	Type	Notes/Example
Address Line	<line>	Internationalized string	As many as are needed
Locality	<locality>	Flexible Property	May be a URI, QCode or Literal value, or no value with a <name> child element
Area	<area>	Flexible Property	
Country	<country>	Flexible Property	
Postal Code	<postalCode>		

For example:

```
<address role="addrole: postal">
  <line>Postfach 16 03 19</line>
  <locality>
    <name>Frankfurt am Main</name>
  </locality>
  <country qcode="ISOCountryCode: de">
    <name xml:lang="en">Germany</name>
  </country>
  <postalCode>D- 60066</postalCode>
</address>
```

12.5.4 Putting it together

The complete concept listing for this example:

LISTING 11 Person Concept conveyed in a NewsML-G2 Concept Item

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20080229:ncdci-person">
```




```

version="1010181123618"
standard="NewsML-G2"
standardversion="2.18"
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<rightsInfo>
  <copyrightHolder>
    <name>IPTC - International Press Telecommunications Council, 20
      Garrick Street, London WC2E 9BT, UK</name>
  </copyrightHolder>
  <copyrightNotice>Copyright 2008, IPTC, www.iptc.org, All
    Rights Reserved</copyrightNotice>
</rightsInfo>
<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2012-11-07T12:38:18Z</versionCreated>
  <firstCreated>2008-12-29T11:00:00Z</firstCreated>
  <pubStatus qcode="stat:usable" />
  <title xml:lang="en">Concept Item describing Mario Draghi</title>
</itemMeta>
<concept>
  <conceptId created="2009-01-10T12:00:00Z" qcode="people:329465" />
  <type qcode="cpnat:person" />
  <name xml:lang="en-GB">Mario Draghi</name>
  <definition xml:lang="en-GB" role="drol:biog">Mario Draghi, born 3 September 1947,
    is an Italian banker and economist who succeeded Jean-Claude Trichet as the
    President of the European Central Bank on 1 November 2011. He was previously
    the governor of the Bank of Italy from January 2006 until October 2011. In 2014
    Forbes nominated Draghi 9th most powerful person in the world.<br />
  </definition>
  <note xml:lang="en-GB" role="nrol:disambiguation">
    Not Mario D'roggia, international powerboat racer
  </note>
  <related rel="relation:occupation" qcode="jobtypes:puboff" />
  <sameAs type="cpnat:person" qcode="pers:567223">
    <name>DRAGHI, Mario</name>
  </sameAs>
  <personDetails>
    <born>1947-09-03</born>
    <affiliation type="orgnat:employer" qcode="org:ECB">
      <name>European Central Bank</name>
    </affiliation>
    <contactInfo role="contactrole:official">
      <email role="addressrole:office">info@ecb.eu</email>
      <im role="imsrvc:reuters">president.ecb.eu@reuters.net</im>
      <phone role="phonerole:switch">+49 69 13 44 0</phone>
      <fax role="faxrole:central">+49 69 13 44 60 00</fax>
      <web>www.ecb.eu</web>
      <address role="addrole:physical">
        <line>Kaiserstrasse 29</line>
        <locality>
          <name>Frankfurt am Main</name>
        </locality>
        <country qcode="ISOCountryCode:de">
          <name xml:lang="en">Germany</name>
        </country>
        <postalCode>D-60311</postalCode>
      </address>
      <address role="addrole:postal">
        <line>Postfach 16 03 19</line>
        <locality>
          <name>Frankfurt am Main</name>
        </locality>
        <country qcode="ISOCountryCode:de">
          <name xml:lang="en">Germany</name>
        </country>
        <postalCode>D-60066</postalCode>
      </address>
    </contactInfo>
  </personDetails>
</concept>
</conceptItem>

```

12.6 More real-world entities

12.6.1 Organisation Details <organisationDetails>

A concept of type “organisation” may hold the following additional properties:

12.6.1.1 Founded <founded> and Dissolved <dissolved>

The date of foundation / dissolution of the organisation, equivalent to born/died for a person, for example

| <founded>1998-06-01</founded>

Or

| <founded>1998</founded>

(See note on Truncated Date Time Property Type in [12.5.3.1](#))

12.6.1.2 Location <location>

A place where the organisation is located, expressed as Flexible Property, NOT an address, repeated as many times as needed. For example:

| <location type=“*loctypes: regoff*” qcode=“*poi: 75001*”>
| <name>Paris</name>
| </location>

12.6.1.3 Contact Information <contactInfo>

Contact information associated with the organisation, uses the same structure as described in [12.5.3.4](#).

12.6.2 Geopolitical Area Details <geoAreaDetails>

A “geoArea” concept may have the following additional properties:

12.6.2.1 Position <position>

This expresses the coordinates of the concept using the following attributes:

Attribute Name	Attribute	Type	Notes/Example
Latitude	@latitude	XML Decimal	The latitude in decimal degrees Positive value = north of the Equator Negative value = south of the Equator
Longitude	@longitude	XML Decimal	The longitude in decimal degrees Positive value = east of the Greenwich Meridian Negative value = west of the Greenwich Meridian
Altitude	@altitude	XML Integer	The absolute altitude in metres with reference to mean sea level
GPS Datum	@gpsdatum	XML String	The GPS datum associated with the position measurement, default is WGS84

12.6.2.2 Founded

The Date and optionally the time plus time zone, that the geopolitical area was founded

| <founded>1998</founded>

12.6.2.3 Dissolved

The Date and optionally the time plus time zone, that the geopolitical area was dissolved.

12.6.3 Point of Interest <poiDetails>

A Point of Interest (POI) is a place “on the map” of interest to people, which is not necessarily a geographical feature, for example concert venue, cinema, sports stadium. As such it has different properties to a purely-geographical point. POI may have the additional properties listed below.

12.6.3.1 Address

The location of the point of interest expressed as a postal address. The <address> element is a wrapper for child elements described in [12.5.3.4](#). In this context, the address is expressly the location of the POI, whereas the <address> wrapper when used as a child of <contactInfo> (see below) expresses the address of the entity who should be contacted about the POI, which could be an office some distance away.

12.6.3.2 Position <position>

The coordinates of the location as described in [12.6.2.1](#)

12.6.3.3 Opening Hours <openHours>

The opening hours of the POI are expressed as a Label type, which is an internationalized string – a natural language expression – extended to include @role if required. Example:

```
<openHours>9.30am to 5.30pm, closed for lunch from 1pm to 2pm</openHours>
```

12.6.3.4 Capacity <capacity>

The capacity of the POI is expressed as a Label:

```
<capacity>10.000 seats</capacity>
```

12.6.3.5 Contact Information <contactInfo>

Contact information for the POI uses the <contactInfo> structure as described in [12.5.3.3](#). It expresses who should be contacted regarding the POI. This could be an organisation located miles away from the location of the POI.

12.6.3.6 Access details <access>

Methods of accessing the POI, including directions. This is a Block type of element, allowing some mark-up and may be repeated as often as needed:

```
<access role="travel type: public">
  The Jubilee Line is recommended as the quickest route to ExCeL London. At Canning
  Town change to the DLR (upstairs on platform 3) for the quick two-stop journey
  to Custom House for ExCeL Station.
</access>
<access role="travel type: road">
  When driving to ExCeL London follow signs for Royal Docks, City Airport and ExCeL
  There is easy access to the M25, M11, A406 and A13.
</access>
```

12.6.3.7 Detailed Information <details>

Detailed information about the location of the POI expressed as a Block type:

```
<details>Room M345, 3rd Floor</details>
```

12.6.3.8 Creation of POI <created>

The date (and optionally a time) on which the Point Of Interest was created.

```
<created>2014-06-23</created>
```

12.6.3.9 Destruction or teardown of POI <ceasedToExist>

The date (and optionally a time) on which the Point Of Interest ceased to exist; perhaps in reference to a temporary POI:

```
<ceasedToExist>2014-06-23</ceasedToExist>
```

12.6.4 Object Details <objectDetails>

Objects that may be expressed as a concept include works of art, books, inventions and industrial artefacts. The IPTC provides three properties for Objects as part of NewsML-G2, but as with any of the types of concept discussed, providers are able to extend the standard for their specific purposes and that of their partners. Note that these are properties of the object described by the Concept, not to be confused with these properties when used in <itemMeta> which apply to the Concept Item conveying the Concept.

The standard additional properties of an Object concept are:

12.6.4.1 Creation Date <created>

The date and, optionally, the time and time zone when the object was created. Non-repeatable.

```
<created>1994-06-14</created>
```

12.6.4.2 Creator <creator>

A party (person or organisation) that created the object, expressed as a Flexibly Property type. Repeatable.

```
<creator type="cpnat:org" qcode="nyse:ba">
  <name>The Boeing Company</name>
</creator>
```

In this case, the object is a Boeing 777 airliner.

12.6.4.3 Copyright Notice <copyrightNotice>

Any necessary copyright notice for claiming the intellectual property of the object. A repeatable Label type:

```
<copyrightNotice role="iprole:company">
  (c) 2008 Boeing Aircraft, all rights reserved
</copyrightNotice>
```

12.7 Relationships between Concepts

This is a group of four properties, <broadener> <narrower> <related> and <sameAs> that enable the creation of particular types of relationship to another concept. For example, our subject was born in Rome. We could create a concept for Rome as follows, with a <broadener> property that denotes that the city as part the region of Lazio:

```
<concept>
  <conceptId qcode="urban:roma" />
  <type qcode="cpnat:geoArea" />
  <definition role="drol:short">Rome (Italian: Roma) is a city and special commune
    (named "Roma Capitale") in Italy. Rome is the capital of Italy and also of the
    homonymous province and of the region of Lazio.<br />
  </definition>
  <broadener type="cpnat:geoArea" qcode="locale:lazio">
    <name xml:lang="en">Lazio (region)</name>
  </broadener>
</concept>
```

<narrower> expresses the reverse relationship. A concept for Rhône could have a <narrower> property linking it to Lyon, and a <broadener> link to the concept of its parent region, or to the concept of the country, France.

<sameAs> allows the provider to inform the recipient that this concept has an equivalent concept in some other taxonomy. For example, we may know that AFP's knowledge base of people has an entry for Mario Draghi that can be referenced using the appropriate alias.

```
<sameAs type="cpnat:person" qcode="AFPpers:567223">
```

```
<name>DRAGHI, Mario</name>
</sameAs>
```

The sameAs property also assists inter-operability because it can be used to enable recipients to choose the CV, or standard, they employ.

For example, the document may have a concept for Germany identified by the provider's QCode "country:de". Some recipients may have standardized on using ISO-3166 Country Codes to classify nationality. The provider can assist recipients to make a direct reference to their preferred scheme using "sameAs":

```
<sameAs qcode="iso3166a2: DE" />
<sameAs qcode="iso3166a3: DEU" />
<sameAs qcode="iso3166n: 276" />
```

<related> allows the expression of a relationship with another concept that cannot be expressed using <broader>, <narrower>, or <sameAs>. For example, the "European Central Bank" may be "related to" "Mario Draghi" – thus the ECB concept may include:

```
<related rel="relation: hasPresident" type="cpnat: person" qcode="people: 329465">
  <name>Mario Draghi</name>
</related>
```

The nature of the relationship is expressed using @rel; the example above indicates that the European Central Bank "has a President" Mario Draghi. This relationship must be part of a CV of relationships, which might include "has a CEO", "has a Finance Director". The IPTC recommends that the <related> property should always contain a @rel.

At PCL, the property may be extended by adding @rank, a numeric ranking of the current concept amongst other concepts related to the target concept.

For example, if the European Central Bank is the second most important concept related to Mario Draghi, amongst other concepts related to him, we can express this as follows:

```
<related rank="2" rel="relation: hasPresident" type="cpnat: person"
  qcode="people: 329465">
  <name>Mario Draghi</name>
</related>
```

12.7.1 Expressing quantitative values using <related>

The <related> property also allows the expression of quantitative values, for example a share price or a sport score, in addition to the concept relationships described above.

The three attributes of <related> that enable this feature are @value, @valueunit and @valuedatatype. Implementers can type the @value by applying an XML Schema datatype, and optionally to declare the units of the value (e.g. the currency) using @valueunit and @valuedatatype.

For example, to express scores in a sports game where the named team won by 4 goals to 2 and gained 3 points:

```
<concept xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <conceptId qcode="ukprem: WBA" />
  <name>West Bromwich Albion</name>
  ...
  <related rel="crel: scoreFor" value="4" valueunit="valueunits: goals"
    valuedatatype="xs: nonNegativeInteger" />
  <related rel="crel: scoreAgainst" value="2" valueunit="valueunits: goals"
    valuedatatype="xs: nonNegativeInteger" />
  <related rel="crel: pointsAdded" value="3" valueunit="valueunits: points"
    valuedatatype="xs: nonNegativeInteger" />
  ...
</concept>
```

A further example, expresses a recommendation from an analyst in the financial markets where the EUR changes from 39 to 44 (expressed as a value), and the rank changes from Hold to Buy (expressed as a value or QCode):

```
<concept xmlns:xs="http://www.w3.org/2001/XMLSchema">
  ...
  <related rel="crel: price_new" value="44" valueunit="iso4217: EUR">
```

```

    <valuedatatype="xs:decimal" />
    <related rel="crel:price_old" value="39" valueunit="iso4217:EUR"
    <valuedatatype="xs:decimal" />
    <related rel="crel:rank_old" value="Hold" valueunit="valueunits:trRanks"
    <valuedatatype="xs:string" />
    <related rel="crel:rank_new" value="Buy" valueunit="valueunits:trRanks"
    <valuedatatype="xs:string" />
    <related rel="crel:rank_new" qcode="trRanks:Buy" />
    ...
</concept>

```



When using <related>: ONLY ONE out of @qcode, @uri, @literal, OR @value MUST be used. (i.e. these properties are mutually exclusive).

The @value has a datatype XML Schema String. If using @value, a @valuedatatype MUST also be used and its datatype must be one of the data types defined by the W3C XML Schema specification. The inclusion of @valueunit is optional, for example if @rel="crel:noOfSiblings" and @value="2" the type of units is obvious.

12.8 Supplementary information about a Concept

Links can be used to enhance the information carried by a NewsML-G2 Concept. For example, a Concept may represent a person in the news; it may also contain some key facts about the person and relationships to other concepts (e.g. membership of an organisation). Links to other resources can also be used to add articles, pictures and other objects to the Concept.

However, the use of <link> as a child of <itemMeta> in a Concept Item would create a problem if a number of Concept Items containing Links were to be aggregated into a Knowledge Item: only the content of the <concept> wrapper would be carried across into the Knowledge Item and the Concept Item Metadata and any Links, would be lost.

To resolve this issue, a <remoteInfo> property may be added to <concept>, with a datatype of LinkType (CCL) and Link1Type (PCL), matching that of <link>. This enables implementers to provide links to supplementary information inside the <concept> wrapper, and thus into a Knowledge Item:

```

<concept>
  <conceptId created="2009-01-10T12:00:00Z" qcode="people:329465" />
  <type qcode="cpnat:person" />
  <name xml:lang="en-GB">Mario Draghi</name>
  <definition xml:lang="en-GB" role="drol:biog"> Mario Draghi, born 3 September
    1947, is an Italian banker and economist who succeeded Jean-Claude Trichet
    as President of the European Central Bank on 1 November 2011.<br />
</definition>
  <related rel="relation:occupation" qcode="jobtypes:puboff" />
  <remoteInfo
    rel="irel:seeAlso"
    contenttype="image/jpeg"
    residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
    <title>
      ECB official portrait picture of Mario Draghi
    </title>
  </remoteInfo>
  <personDetails>
    <born>1942-12-20</born>
    ...
    <contactInfo role="contactrole:official">
    ...
    </contactInfo>
  </personDetails>
</concept>

```

Annotations in the original image:

- Arrow from "link" start to the start of <remoteInfo>
- Arrow from "link" end to the end of </remoteInfo>
- Arrow from "Item ref" to the residref attribute value.

Using the rules given in 26.2.2.4, when adding properties of the target NewsML-G2 Item, the parent property must be included if it is not from either <contentMeta> or <itemMeta>. For example, a <description> element extracted from <contentMeta> (no parent needed):

```

<remoteInfo
  rel="irel:seeAlso"
  contenttype="video/mpeg"
  residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
  <description>
    ECB official video of Mario Draghi working with senior
    colleagues at the Bank
  </description>
</remoteInfo>

```




| </descri pti on>

Contrast with a <description> from <partMeta>, which must be included as the parent element:

| </remoteInfo>
| <partMeta partid="part1" seq="1"
| <description>The first part shows...</descri pti on>
| </partMeta>
| <partMeta partid="part2" seq="2"
| <description>The second part shows...</descri pti on>
| </partMeta>
| </remoteInfo>

12.9 Concepts in Practice

The more common method of exchanging Concepts is as part of a Controlled Vocabulary (aka Taxonomy, Thesaurus, Dictionary, etc), which are conveyed in NewsML-G2 as a set of concepts in a Knowledge Item. This is discussed in [Knowledge Items](#).

The use of Concepts to convey Event information is discussed in [EventsML-G2](#).



This page intentionally blank

13 Knowledge Items

13.1 Introduction

When news happens, the events rarely take place in isolation. There will be a series of relationships between the news event and the people, places and organisations that are directly or indirectly involved. Many of these entities will be well-known, and readers of the news may expect to be able to navigate to further information about these entities, or to find other events in which they are involved. This expectation is heavily fostered by people's familiarity with the Web: they expect to be able to click on the name of, say, a company and see more information about it.

There will also be references to abstract notions such as subject classifications that enable the news event to be searched and sorted according to user preferences.

To fully exploit the value of their services, news organisations need to be able to exchange this supporting information in an industry-standard way that can be processed using standard technology.

As described in the previous chapter **Concepts and Concept Items**, NewsML-G2 has powerful features for encapsulating this detailed information about entities and notions, but the Concept Item can convey only a single concept, while the Knowledge Item is able to convey many of them, even a full taxonomy.

For example, a profile of, say, Russian President Vladimir Putin can be conveyed in a Concept. By including this Concept in a set of similar Concepts profiling world leaders, we can create and exchange a Knowledge Base of political personalities that can be updated and referred to over time.

The level of detail of information that a provider may make available in a Knowledge Item will depend on its business model and relationship with the receiving customer(s). Providers may make variable levels of information available according to subscription since it is clear that the content of their Knowledge Bases is likely to be valuable.

There are also opportunities for third-party providers of specialist information to partner with providers and customers to create value-added knowledge services using a NewsML-G2 infrastructure.

IPTC NewsCodes are controlled vocabularies maintained as Knowledge Items and are available on the IPTC web site at <http://www.iptc.org/newscodes/>. Choose View NewsCodes and follow instructions for downloading any of the CVs.

13.2 Example: a Knowledge Item for <accessStatus>

One of the available properties for describing a news event is <accessStatus> to provide information about the physical accessibility of the place where an event is due to occur.

The property takes a QCode value:

```
<conceptId qcode="access: restricted" />
```

As there is no IPTC NewsCodes scheme currently defined for Access Status, any provider wishing to include this information would need to create a Controlled Vocabulary.

This example shows a Knowledge Item which defines a Controlled Vocabulary of Access Status terms. This CV would be available as a Knowledge Item, with names and definitions in English, French and German, as shown below:

Value	Names	Definitions
easy	Easy access	Unrestricted access for vehicles and equipment. Loading bays and/or lifts for unimpeded access to all levels
	Facile d'accès	Un accès sans restriction pour les véhicules et l'équipement. Les quais de chargement et / ou des ascenseurs pour l'accès sans entrave à tous les niveaux

Value	Names	Definitions
	Der Zugang ist einfach	Ungehinderten Zugang für Fahrzeuge und Ausrüstung. Laderampen und / oder Aufzüge für den uneingeschränkten Zugang zu allen Ebenen
restricted	Access is Restricted Accès Restreint	Access for vehicles and equipment possible but restricted. There may be obstacles, height or width restrictions that will impede large or heavy items. Advise checking with the organizers. L'accès des véhicules et de matériel possible, mais limitée. Il y mai être des obstacles, la hauteur ou la largeur des restrictions qui empêchent les grandes ou d'objets lourds. Conseiller à la vérification avec les organisateurs.
	Der Zugang ist eingeschränkt	Zugang für Fahrzeuge und Ausrüstung möglich, aber eingeschränkt. Möglicherweise gibt es Hindernisse, Höhe und Breite, die Beschränkungen behindern große oder schwere Gegenstände. Beraten Sie mit dem Veranstalter in Verbindung
difficult	Access is difficult L'accès est difficile	Access includes stairways with no lift or ramp available. It will not be possible to install bulky or heavy equipment that cannot be safely carried by one person Comprend l'accès aux escaliers ou la rampe sans ascenseur disponible. Il ne sera pas possible d'installer des équipements lourds ou volumineux qui ne peuvent pas être transportés en sécurité par une seule personne
	Der Zugang ist schwierig	Access enthält Treppen ohne Fahrstuhl oder Rampe zur Verfügung. Es wird nicht möglich sein, installieren sperrige oder schwere Geräte, die sich nicht sicher befördert werden von einer Person

13.3 Structure and Properties

Knowledge Items share a common structure with News Items, Package Items and Concept Items.

This Chapter assumes that the reader is familiar with the chapter on [Concepts and Concept Items](#).

13.3.1 The <knowledgeItem> element

The top level element of a Knowledge Item is <knowledgeItem>, which contains id, versioning and catalog information.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<knowledgeItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20090202:ncdki-accesscode"
  version="20141018"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power" >
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
```

13.3.2 Item Metadata

The <itemMeta> block contains management metadata for the Knowledge Item document. Below is a minimum set of properties.

The Item Class property should use the IPTC "Nature of Concept Item" NewsCode (scheme alias "cinat"). The appropriate value in the case of sending a CV or taxonomy is "scheme", denoting that this is a full scheme of concepts contained in this Knowledge Item.

```
<itemMeta>
  <itemClass qcode="cinat:scheme" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2014-11-08T00:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
```

13.3.3 Content Metadata

The optional <contentMeta> block contains Administrative Metadata and Descriptive Metadata shared by the concepts conveyed by the <conceptSet>.

13.3.3.1 Administrative Metadata

This example timestamps the content:

```
<contentMeta>
  <contentModified>2009-01-28T13:00:00Z</contentModified>
  ....
```

More details about informing receivers about changes to Knowledge Item content are contained in

14.7.2.4 Handling updates to Knowledge Items using

13.3.3.2 Descriptive Metadata

The descriptive metadata properties <subject> and <description> may be used by Knowledge items, in any order. They are optional and repeatable. This example uses the <description> element:

```
<description xml:lang="en">
  Classification of the ease of gaining physical access to the location of a news
  event for the purpose of deploying personnel, vehicles and equipment.
</description>
<description xml:lang="fr">
  Classification de la facilité d'obtenir un accès physique à l'emplacement d'un
  événement pour le déploiement de personnel, de véhicules et d'équipements.
</description>
<description xml:lang="de">
  Klassifikation der physischen Zugriff auf den Standort eines News Termine für
  Die Zwecke der Bereitstellung von Personal, Fahrzeugen und Ausrüstungen.
</description>
```

13.3.4 Concept Set

A single <conceptSet> element wraps zero or more <concept> components. The order of the Concepts is not important. Properties of <concept> are optional and described in [Concepts and Concept Items](#).

Each member of the CV requires its own <concept> wrapper with a Concept ID and Name within the Concept Set:

```
<conceptSet>
  <concept>
    <conceptId qcode="access:easy" />
    <name xml:lang="en">Easy access</name>
    ....
  </concept>
  <concept>
    <conceptId qcode="access:difficult" />
    <name xml:lang="en">Access is difficult</name>
    ....
  </concept>
  <concept>
    <conceptId qcode="access:restricted" />
    <name xml:lang="en">Access is Restricted</name>
    ....
  </concept>
</conceptSet>
```

Each Concept also has a <definition> in three languages that gives further details in natural language, for example the English definition:

```
<definition xml:lang="en">
  Access for vehicles and equipment possible but restricted. There may be
  obstacles, height or width restrictions that will impede large or heavy
```

```

    items. Advise checking with the organisers.
</definition>

```

This completes the first Concept in the <conceptSet>. The two other concepts in the CV are added in a similar fashion.

13.3.5 Scheme Metadata

In NewsML-G2 v 2.17, the <schemeMeta> element was added to enable a Knowledge Item to support similar properties to the <scheme> in a Catalog. It should be noted that <schemeMeta> is used to express metadata about the scheme being conveyed, and should ONLY be used where the Knowledge Item contains all the concepts from a *single* scheme, as denoted by:

```
<itemClass qcode="cinat:scheme"/>
```

The <schemeMeta> element is used after <conceptSet> and contains the same attributes and child properties of <scheme> in a catalog, except for @alias, but additionally has a <related> and a <schemeMetaExtProperty> child element. Also new are the attributes @authority, to indicate the party who controls the scheme, and @preferredalias, to enable scheme authority to express the recommended scheme alias to be used with QCodes.

One of the benefits of this approach is that it enables implementers to express the top-level concept(s) of a scheme, a requirement of SKOS.

```

<schemeMeta uri="http://cv.example.org/newscodes/access/"
authority="http://www.example.org" preferredalias="access">
  <definition xml:lang="en-GB">Classification of the ease of gaining physical access
to the location of a news event for the purpose of deploying personnel, vehicles
and equipment.</definition>
  <name xml:lang="en-GB">Ease of Access</name>
  <related qcode="access:easy" rel="skos:hasTopConcept"/>
  <related qcode="access:difficult" rel="skos:hasTopConcept"/>
  <related qcode="access:restricted" rel="skos:hasTopConcept"/>
</schemeMeta>

```

LISTING 12 Knowledge Item for Access Codes

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<knowledgeItem xmlns="http://iptc.org/std/nar/2006-10-01/"
guid="urn:newsml:iptc.org:20090202:ncdki-accesscode" version="20101018"
standard="NewsML-G2"
standardversion="2.18"
conformance="power">
  <catalogRef
href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml"/>
  <itemMeta>
    <itemClass qcode="cinat:scheme"/>
    <provider qcode="nprov:IPTC"/>
    <versionCreated>2014-11-08T00:00:00Z</versionCreated>
    <pubStatus qcode="stat:usable"/>
  </itemMeta>
  <contentMeta>
    <contentCreated>2009-01-28T12:00:00Z</contentCreated>
    <contentModified>2009-01-28T13:00:00Z</contentModified>
    <description xml:lang="en">
      Classification of the ease of gaining physical access to the location of a news
      event for the purpose of deploying personnel, vehicles and equipment.
    </description>
    <description xml:lang="fr">
      Classification de la facilité d'obtenir un accès physique à l'emplacement d'un
      événement pour le déploiement de personnel, de véhicules et d'équipements.
    </description>
    <description xml:lang="de">
      Klassifikation der physischen Zugriff auf den Standort eines News Termine für
      Die Zwecke der Bereitstellung von Personal, Fahrzeugen und Ausrüstungen.
    </description>
  </contentMeta>
  <conceptSet>
    <concept>
      <conceptId qcode="access:easy"/>
      <name xml:lang="en">Easy access</name>
      <name xml:lang="fr">Facile d'accès</name>
      <name xml:lang="de">Der Zugang ist einfach</name>
      <definition xml:lang="en">

```

```

    Unrestricted access for vehicles and equipment. Loading bays
    and/or lifts for unimpeded access to all levels
</definition>
<definition xml:lang="fr">
    Un accès sans restriction pour les véhicules et l'équipement. Les quais de
    chargement et / ou des ascenseurs pour l'accès sans entrave à tous les
    niveaux
</definition>
<definition xml:lang="de">
    Ungehinderten Zugang für Fahrzeuge und Ausrüstung. Laderampen und / oder
    Aufzüge für den uneingeschränkten Zugang zu allen Ebenen
</definition>
</concept>
<concept>
    <conceptId qcode="access: difficult" />
    <name xml:lang="en">Access is difficult</name>
    <name xml:lang="fr">L'accès est difficile</name>
    <name xml:lang="de">Der Zugang ist schwierig</name>
    <definition xml:lang="en">
        Access includes stairways with no lift or ramp available. It will not be
        possible to install bulky or heavy equipment that cannot be safely carried
        by one person
    </definition>
    <definition xml:lang="fr">
        Comprend l'accès aux escaliers ou la rampe sans ascenseur disponible. Il ne
        sera pas possible d'installer des équipements lourds ou volumineux qui ne
        peuvent pas être transportés en sécurité par une seule personne
    </definition>
    <definition xml:lang="de">
        Access enthält Treppen ohne Fahrstuhl oder Rampe zur Verfügung. Es wird nicht
        möglich sein, installieren sperrige oder schwere Geräte, die sich nicht
        sicher befördert werden von einer Person
    </definition>
</concept>
<concept>
    <conceptId qcode="access: restricted" />
    <name xml:lang="en">Access is Restricted</name>
    <name xml:lang="fr">Accès Restreint</name>
    <name xml:lang="de">Der Zugang ist eingeschränkt</name>
    <definition xml:lang="en">
        Access for vehicles and equipment possible but restricted. There may be
        obstacles, height or width restrictions that will impede large or heavy
        items. Advise checking with the organisers.
    </definition>
    <definition xml:lang="fr">
        L'accès des véhicules et de matériel possible, mais limitée. Il y mai être
        des obstacles, la hauteur ou la largeur des restrictions qui empêchent les
        grandes ou d'objets lourds. Conseiller à la vérification avec les
        organiateurs.
    </definition>
    <definition xml:lang="de">
        Zugang für Fahrzeuge und Ausrüstung möglich, aber eingeschränkt.
        Möglicherweise gibt es Hindernisse, Höhe und Breite, die Beschränkungen
        behindern große oder schwere Gegenstände. Beraten Sie mit dem Veranstalter in
        Verbindung
    </definition>
</concept>
</conceptSet>
<schemeMeta uri="http://cv.example.org/newscodes/access/"
    authority="http://www.example.org" preferredalias="access">
    <definition xml:lang="en">Classification of the ease of gaining physical access
    to the location of a news event for the purpose of deploying personnel,
    vehicles and equipment.</definition>
    <name xml:lang="en-GB">Ease of Access</name>
    <related qcode="access: easy" rel="skos:hasTopConcept"/>
    <related qcode="access: difficult" rel="skos:hasTopConcept"/>
    <related qcode="access: restricted" rel="skos:hasTopConcept"/>
</schemeMeta>
</knowledgeItem>

```

13.4 Knowledge Workflow

Figure 12 shows a possible information flow for news information that exploits the possibilities of NewsML-G2 Concepts and Knowledge Items to add value to news:

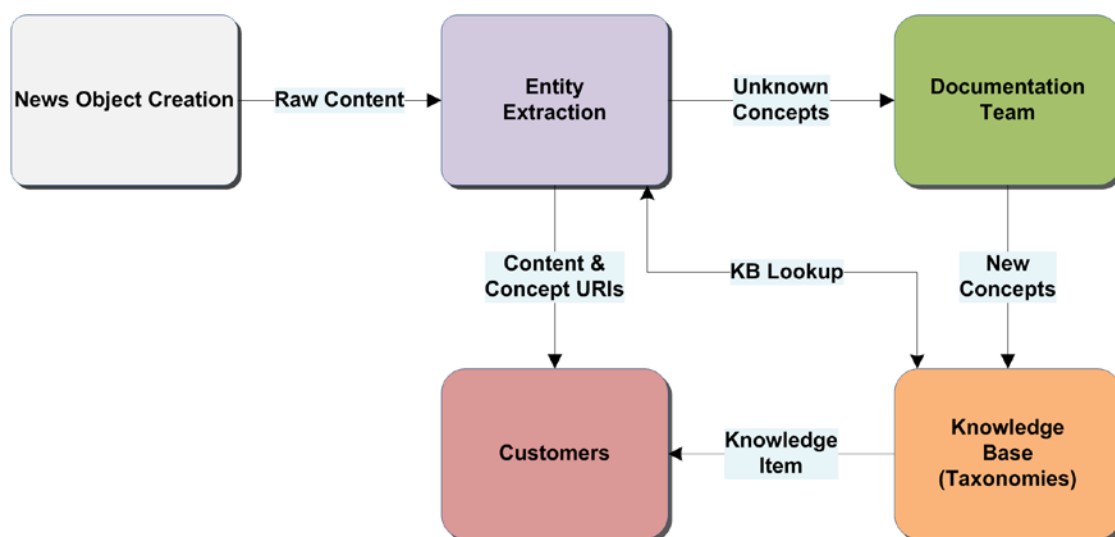


Figure 12: Information Flow for Concepts and Knowledge Items

Increasingly, news organisations are using entity extraction engines to find “things” mentioned in news objects. The results of these automated processes may be checked and refined by journalists. The goal is to classify news as richly as possible and to identify people, organisations, places and other entities before sending it to customers, in order to increase its value and usefulness.

This entity extraction process will throw up exceptions – unrecognised and potentially new concepts – that may need to be added to the Knowledge Base. Some news organisations have dedicated documentation departments to research new concepts and maintain the Knowledge Base.

When new concepts are submitted to the Knowledge Base, they are added to the appropriate taxonomy and may be made available to customers (depending on the business model adopted) either partially or fully as Knowledge Items.

13.5 Using G2 Knowledge Items with SKOS

The Simple Knowledge Organisation System is a W3C standard for using RDF-based means to share information about knowledge organization systems (see www.w3.org/2004/02/skos/).

The IPTC Media Topic NewsCodes are a working example of how a G2 Knowledge Item may have features added that can align a G2 Scheme to SKOS using the <related> child element of <concept> and <schemeMeta>.

For example, the Media Topic for the sport of biathlon, part of the Media Topics scheme (<http://cv.iptc.org/newscodes/mediatopic/>):

```

<concept id="medtop20000852" modified="2010-12-14T21:53:19+00:00">
  <conceptId qcode="medtop:20000852" created="2009-10-22T02:00:00+00:00"/>
  <type qcode="cpnat:abstract"/>
  <name xml:lang="en-GB">biathlon</name>
  <definition xml:lang="en-GB">A combination of cross country skiing and target shooting on a 12.5 K course in a pursuit format. </definition>
  <broadener qcode="medtop:20000822"/>
  <related qcode="medtop:20000822" rel="skos:broadener"/>
  <related qcode="subj:15009000" rel="skos:exactMatch"/>
  <related uri="http://cv.iptc.org/newscodes/mediatopic/" rel="skos:inScheme"/>
</concept>
  
```

This uses the G2 <broadener> property to express that the “biathlon” topic (medtop:20000852) is a child of the “competition discipline” topic (medtop:20000822), and this is complemented by the using a <related> to indicate the same relationship by the SKOS term “skos:broadener”. The second <related> element indicates that the legacy IPTC Subject Code NewsCode 15009000 is the exact match (G2 = <sameAs>) for this Media Topic.

The CV also contains a Scheme Metadata element as follows:

```

<schemeMeta uri="http://cv.iptc.org/newscodes/mediatopic/"
  authority="http://www.iptc.org" preferredalias="medtop">
  
```




```
<definition xml:lang="en-GB">Indicates a subject of an item.</definition>
<name xml:lang="en-GB">Media Topic</name>
<note xml:lang="en-GB">The Media Topic NewsCodes is IPTC's new (as of December
  2010) 1100-term taxonomy with a focus on text. The development started with the
  Subject Codes and extended the tree to 5 levels and reused the same 17 top
  level terms. The terms below the top level have been revised and rearranged.
  Each Media Topic provides a mapping back to one of the Subject Codes. </note>
<related qcode="medtop:01000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:02000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:03000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:04000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:05000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:06000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:07000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:08000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:09000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:10000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:11000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:12000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:13000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:14000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:15000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:16000000" rel="skos:hasTopConcept"/>
<related qcode="medtop:17000000" rel="skos:hasTopConcept"/>
</schemeMeta>
```

This satisfies the SKOS requirement to indicate the top-level concepts of any scheme.

The current structure of G2 Knowledge Items delivered by IPTC is documented on this web page:
<http://dev.iptc.org/NewsCodes-G2-Knowledge-Items-by-IPTC>



This page intentionally blank

14 Controlled Vocabularies and QCodes

14.1 Introduction

One of the fundamental ideas underpinning NewsML-G2 is the use of Controlled Vocabularies (CVs) or taxonomies to enable two basic operations:

- ❖ To restrict the allowed values of certain properties in order to maintain the consistency and interoperability of machine-readable information – supplying data to populate menus, for example.
- ❖ To provide a concise method of unambiguously identifying any abstract notion (e.g. subject classification) or real-world entity (person, organisation, place etc.) present in, or associated with, an item. This enables links to be made to external resources that can provide the consumer with further information or processing options.

A Controlled Vocabulary (CV) is a set of concepts usually controlled by an authority which is responsible for its maintenance, i.e. adding and removing vocabulary entries. In NewsML-G2, CVs are also known as Schemes. The person or organisation responsible for maintaining a Scheme is the Scheme Authority.

Examples of CVs include the set of country codes maintained by the International Standards Organisation, and the NewsCodes maintained by the IPTC. An application of a CV could be a drop-down list of countries in an application interface.

Many CVs are dedicated to a specific metadata property, for example there are CVs for <subject>, for <genre>; or they are dedicated to a specific attribute that refines a property e.g. CV for the @role of <description>.

In news distributions that use NewsML-G2, it is recommended that Controlled Vocabularies are exchanged as Knowledge Items, with members of the CV contained in individual <concept> structures.

Members of a Scheme are each identified by a concept identifier expressed as a **QCode** (note capitalisation), which is resolved via the Catalog information in the NewsML-G2 Item to form a URI that is globally unique.

14.2 Business Case

Controlled Vocabularies are needed in information exchange because they establish a common ground for understanding content that is language-independent. Schemes and QCodes enable CVs to be exchanged and referenced using Web technology, and provides a lightweight, flexible and reliable model for sharing concepts and information about concepts

For example, the IPTC Media Topics Scheme is a language-independent taxonomy for classifying the subject matter of news. A consumer receiving news classified using this scheme can discover the meaning of this classification, using a publicly-accessible URL.

Examples abound of non-IPTC CVs in everyday use: IANA MIME types, ISO Country Codes or ISO Currency Codes to name but three.

News providers use CVs to add value to their content:

- ❖ News can be accurately processed by software if it adheres to known (i.e. controlled) parameters expressed as a CV, for example the publishing status of a news item
- ❖ by establishing CVs of people, places and organisations, the identity of entities in the news can be unambiguously affirmed;
- ❖ CVs can be extended to store further information about entities in the news, for example biographies of people, contact details for organisations.

14.3 How QCodes work

A QCode is a string with three parts, all of which **MUST** be present:

- ❖ Schema Alias: the prefix, for example “stat”
- ❖ Scheme-Code Separator: a separator, which **MUST** be a colon “:” (ASCII 58_{dec} = 003A_{hex})
- ❖ Code Value: the suffix, for example “usable”



This produces a complete QCode of “stat:usable” represented in XML as

```
| qcode="stat: usable"
```

14.3.1 Scheme Alias to Scheme URI

The key to resolving scheme aliases is the Catalog information, a child of the root element of every NewsML-G2 Item. A scheme alias may be resolved directly using the <catalog> element:

```
| <catalog>  
|   <scheme alias="stat" uri="http://cv.iptc.org/newscodes/pubstatusg2/" />  
| </catalog>
```

Using the information in <catalog>, a processor now has a *Scheme URI* that can be used as the next step in resolving the QCode.

14.3.1.1 Catalogs

The <catalog> element may contain many <scheme> components. Catalog information can be stored in one of two ways:

- ❖ directly in a NewsML-G2 Item using the <catalog> element, or
- ❖ remotely in a file containing the catalog information, referenced by the @href of a <catalogRef> element.

There is likely to be more than one <catalogRef> in an Item:

```
| <catalogRef  
|   href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />  
| <catalogRef  
|   href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
```

These remote catalogs are hosted by specific authorities, in this case by the IPTC, and by the information provider XML Team. Each remote catalog file will contain a <catalog> element and a series of <scheme> components that map the scheme aliases used in the item to their scheme URIs. For a more detailed description of managing Catalogs, see **14.6 Creating and Managing Catalogs** below.

14.3.2 Concept URI

Appending the QCode Code Value to the Scheme URI produces the *Concept URI*.

<http://cv.iptc.org/newscodes/pubstatusg2/usable>

The IPTC recommends that Scheme/Concept URIs can be resolved to a Web resource that contains information in both machine-readable and human-readable form (This is also a recommendation for the Semantic Web), i.e. they are URLs. The concept resolution mechanism used by the IPTC is http-based, and the NewsML-G2 Specification describes how an http-URL should be resolved.

Entering the above Concept URI in a web browser results in the following page being displayed:

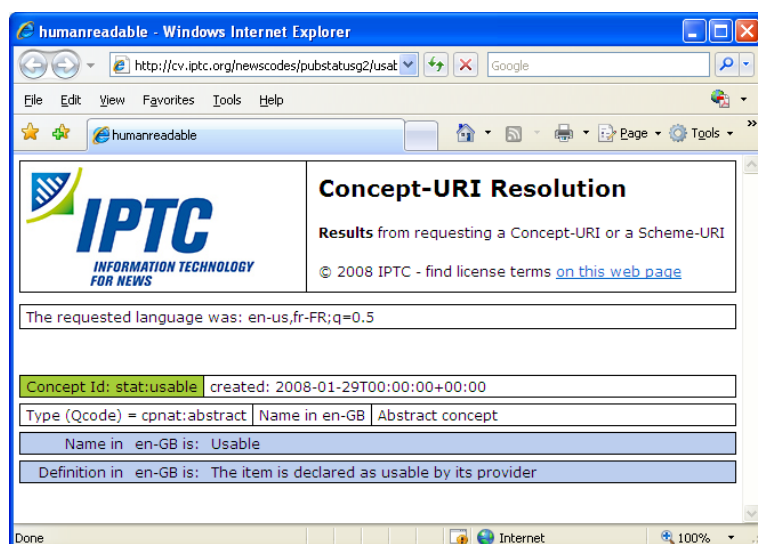


Figure 13: Human-readable browser page of a Concept URI

A question often asked by implementers is: "What happens if I receive files from two providers who inadvertently have a clash of scheme aliases?"

The scenarios they envisage are either:

- ❖ Provider A and Provider B use the same scheme alias to represent different schemes. For example the alias "pers" is used by both providers to represent their own proprietary CVs of people, or
- ❖ Provider A and Provider B use a different scheme alias to represent the same scheme. For example, A uses "subj" to represent the IPTC Subject NewsCodes, and B uses "tema" to represent the same CV.

The answer is "everything works fine!"; QCode to Concept URI mappings must be unique only within the scope of each document in which they appear.

A processor should correctly process two files with different aliases to the same Concept URI:

```
<!-- First Document - scheme alias "subj" -->
<catalog>
  <scheme alias="subj" uri="http://cv.iptc.org/newscodes/subjectcode/" />
  ...
</catalog>
<subject type="cpnat:abstract" qcode="subj:1500000" />
```

```
<!-- Second Document - scheme alias "tema" -->
<catalog>
  <scheme alias="tema" uri="http://cv.iptc.org/newscodes/subjectcode/" />
  ...
</catalog>
<subject type="cpnat:abstract" qcode="tema:1500000" />
```

This is because the concept resolution process is local to each document. The processor can unambiguously resolve the QCode to a Concept URI via the <catalog> in each case.

The following example does NOT work because the same alias is mapped to two different URIs within the same document and the processor is unable to resolve the QCode to a single Concept URI:

```
<catalog>
  <scheme alias="subject" uri="http://cv.iptc.org/newscodes/subjectcode/" />
  ...
  <scheme alias="subject" uri="http://cv.example.com/subjectcodes/codelist/" />
</catalog>
...
<subject type="cpnat:abstract" qcode="subject:1500000" />
...
```

But the following is CORRECT because it is possible to have different aliases within the same document pointing to the same URI and the processor can resolve both QCodes:



```
<catalog>
  <scheme alias="subject" uri="http://cv.iptc.org/newscodes/subjectcode/" />
  ...
  <scheme alias="subj" uri="http://cv.iptc.org/newscodes/subjectcode/" />
</catalog>
...
<subject type="cpnat:abstract" qcode="subject:1500000" />
...
<subject type="cpnat:abstract" qcode="subj:1500000" />
```

In this document, there are many references to IPTC NewsCodes and their scheme aliases. From the above, it will be obvious that these specific alias values are not mandatory, although the IPTC recommends the consistent use of scheme alias values by implementers.

14.4 QCodes and Taxonomies

Taxonomies, also known as thesauri, knowledge bases etc., are repositories of information about notions or ideas, and about real-world “things” such as people, companies and places.

For example, a processor might encounter the following XML in a NewsML-G2 document:

```
<subject type="cpnat:person" qcode="pol:rur12345">
```

The subject property shown here has two QCodes, one for @type, and the other as @qcode. The “cpnat” alias is for a controlled vocabulary of allowed categories of concept, which includes values of “person”, “organisation”, “POI” (point of interest). Using @type in this way enables further processing such as “find all of the people identified in the document”.

The second QCode encountered in the subject is “pol:rur12345”. Resolving this (fictional) scheme alias and suffix might result in the following concept URI:

```
http://www.example.com/knowledgebase/people/political\_leaders/rur12345
```

Fetching the information at the above resource, may return the following information:

PUTIN, Vladimir – Prime Minister of the Russian Federation	
Name (FAMILY, Given)	PUTIN, Vladi mi r Vladi mi rovi tch
Name (known as)	Vladi mi r Vladi mi rovi tch Put in
Summary	Former Soviet intelligence officer who has served terms as Russian president and prime minister.
Background	Born 7 October 1952
Place of birth	Leningrad (St Petersburg)
Other	German- speaker

The IPTC recommends that providers should make schemes containing concepts such as the above available to recipients as Knowledge Items. It should include at least a name; the amount of further knowledge about the concepts could be different for different customer classes and depend on contracts.

14.5 Managing Controlled Vocabularies as NewsML-G2 Schemes

14.5.1 Knowledge Items

In a workflow where partners are exchanging news information using NewsML-G2, Knowledge Items are the most compliant method of distributing Controlled Vocabularies. The sections below describe the steps to create a new CV: first by creating a Scheme (14.5.2) and next creating a Knowledge Item from a set of existing Concept Items (14.5.3) for distribution to customers and partners.

Knowledge Items do not necessarily contain all of the information that a provider possesses about any given set of concepts. This, after all, may be commercially valuable information that the provider makes available on a per-subscriber basis. For example, a lower fee might entitle the subscriber to basic information about a concept, say a person, while a higher fee might give access to full biographical details and pictures.

It is not mandatory that information about CVs be stored or distributed in the technical format of a Knowledge Item. It is sufficient, for the correct processing of a NewsML-G2 Item, only that a Scheme Alias/Code Value pair (defining the Concept URI) is unambiguous. The IPTC makes the following recommendations about CVs:

- ❖ Knowledge Items **SHOULD** be used to distribute CVs. Other means such as paper, fax or email are permissible but at the price of less efficient automated processing.
- ❖ Concept URIs **SHOULD** resolve to a Web resource; this is a requirement for the Semantic Web.
- ❖ In the case where a Scheme Authority does not make the concepts of a CV available as a Web resource, the Scheme URI **SHOULD** resolve to a Web resource, such as a human-readable Web page giving information about the purpose of the CV, and where details of the Scheme can be obtained.

14.5.2 Creating a new Scheme

A NewsML-G2 controlled vocabulary is a set of concepts. To create a CV as a NewsML-G2 Scheme:

- ❖ Assign a Scheme URI which must be an http URL for the Semantic Web (example: `http://cv.example.org/schemeA/`) and a Scheme Alias (example: `abc`)

Add this Scheme Alias and URI to the catalog:

```
<catalog>
  <scheme alias="abc" uri="http://cv.example.org/schemeA/" />
</catalog>
```

If using a remote catalog, change the catalog URI to reflect a new version of the catalog (so that recipients know that they should add this to their cache of catalogs) and ensure that all NewsML-G2 Items using the new Scheme refer to the new version of the remote catalog.

- ❖ Create Concepts as required. You must use the Scheme Alias of the new scheme with the identifier of this new concept. For example:

```
<concept>
<conceptId created="2009-09-22" qcode="abc:concept-x"
...

```

The above Concept Identifier resolves to the Concept URI
`"http://cv.example.org/schemeA/concept-x"`

14.5.3 Creating a new Knowledge Item for distributing as a CV

Chapter 13 on Knowledge Items shows an example Controlled Vocabulary expressed in XML. A Knowledge Item contains concepts from one or more Schemes. The steps to begin creating a KI are:

- ❖ Identify the set of Concept Items that contain the concepts that will be part of the Knowledge Item, which may be only from this new CV or also from other CVs.
- ❖ Create the metadata properties for the Knowledge Item that express the rules used to create it, for example, a `<title>` and `<description>` such as "Concepts extracted from Schemes A and B based on criteria X and Y".
- ❖ For each Concept: Copy all or part of the selected concept details (the `<concept>` wrapper and associated properties) into the Knowledge Item.

14.5.4 Managing Schemes

14.5.4.1 Changes to Schemes

Scheme URIs **MUST** persist over time, and any changes to a Scheme which involve the creation or deprecation of concepts **MUST** be backwardly compatible with existing concepts. (For example, the code of a retired Concept **MUST NOT** be reused for a new and different concept.)

Scheme Authorities can indicate that a member of a CV should no longer be applied as a new value. This must be expressed by adding a `@retired` attribute to the `<conceptId>` of the Concept that is no longer to be used.



Both @created and @retired attributes are of datatype Date with optional Time and Time Zone (DateOptTime) and their use is optional. The @retired date can be a date in the future when a Scheme Authority knows that the Concept ID should no longer be used for new NewsML-G2 Items.

Example of a retired concept:

```
<conceptId created="2006-09-01" retired="2009-12-31" qcode="foo:bar" />
```

- ❖ Concepts MUST NOT be deleted from a Scheme; this could cause processing errors for NewsML-G2 Items that pre-date the changes. Use of @retired ensures that Items that pre-date a CV change will continue to correctly resolve "legacy" concept identifiers.
- ❖ For the same reason, Concept IDs MUST NOT be re-cycled, i.e. the same identifier MUST NOT be used for a different concept.
- ❖ Schemes themselves MUST NOT be deleted, as archived content is likely to use the concepts contained in a retired CV.

14.5.4.2 Recommendations for non-complying Schemes

Some Scheme Authorities may fail to comply with the NewsML-G2 Specification, and this could be beyond the control of the end-user or the provider. Guidelines for handling various scenarios are:

1. *The authority of the vocabulary governs the scheme URI and the code – but does not comply with the G2 Specification*
Reusing a Concept URI which was assigned to one concept with another concept is a breach of the NewsML-G2 specifications. If there are requirements that drive the authority to do so, the authority should give a clear and warning notification about that fact to all receivers at the time of the publication of the reused Concept URI.
2. *The authority of the vocabulary governs only the code but not the scheme URI*
This may be the case for Controlled Vocabularies of codes only and if a news provider assigns a scheme URI of its own domain to enable the CV to be used with NewsML-G2. A good example are the scheme URIs defined by the IPTC for the ISO vocabularies for country names or currencies – see <http://cvx.iptc.org>
The party who assigned the scheme URI has the responsibility of making users of this scheme URI together with the vocabulary aware of any reuse cases – and should post a generic warning about the potential threat of reused codes.
In addition the party who assigned the scheme URI may consider changing the scheme URI in the case of the reuse of a code. This would avoid having the same Concept URI for different concepts but would require careful management of the vocabularies as actually a completely new Controlled Vocabulary is created by the use of the new scheme URI.
3. *Another code is assigned to the same concept or a very likely similar concept*
This use case does not violate the NewsML-G2 specifications. But care should be taken for establishing relationships between Concept URIs:
 - If a new code is assigned to the same concept a sameAs or a SKOS exactMatch relationship of this new URI should be established pointing to all other existing URIs identifying this concept.
 - If a slightly modified concept is created and gets a new Concept URI it may be considered to establish a closeMatch relationship from SKOS [<http://www.w3.org/TR/skos-reference/#mapping>]

14.6 Creating and Managing Catalogs

As previously described, Catalogs are essential to the resolution of NewsML-G2 Controlled Vocabularies and constituent QCodes.

14.6.1 The <catalog> element

Each QCode Scheme Alias that is used in a NewsML-G2 Item must have a reference in the Item's <catalog> to the Controlled Vocabulary in which it is included. A <catalog> contains the Scheme Alias and the Scheme URI:

```
<catalog>
  <scheme alias="timeunit" uri="http://cv.iptc.org/newscodes/timeunit/" />
</catalog>
```

14.6.2 Remote Catalogs

As the CVs used by a provider are usually quite consistent across the NewsML-G2 Items they publish, the IPTC recommends that the <catalog> references are aggregated into a stand-alone file which is made available as a web resource: known as a Remote Catalog. This file is referenced by a <catalogRef> in the Item:

```
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
```

Note: the <catalog> element in such a stand-alone file needs an XML namespace definition:

```
<catalog xmlns="http://iptc.org/std/nar/2006-10-01/">
```

The use of stand-alone web resources is preferable because all of the QCode mappings are shared across many G2 Items; the local <catalog> can only be used by the single Item.

14.6.3 Managing Catalog files

Simple management of Remote Catalogs over time is relatively straightforward: whenever a new scheme is added or the alias or the uri of any of the existing schemes are changed, a new Catalog must be published with a new URL. This URL may reflect a version number of the catalog. This is the method that the IPTC uses to maintain the NewsCodes, simply by increasing the file suffix digit by one:

```
href="..Standards_21.xml" --> href="..Standards_25.xml"
```

Note that the "old" Remote Catalogs must continue to be available as a web resource, otherwise existing G2 Items and QCodes that reference them will not be able to resolve related Scheme Aliases to Scheme URIs.



Receiving applications **MUST** use the catalog information contained in the NewsML-G2 document being processed. If a provider updates a catalog, this is likely to be because new schemes have been added. Using a catalog other than that indicated in the document could cause errors or unintended results.

14.6.3.1 Additional <catalog> features

To improve the management features of Catalogs, new (optional) properties were added to the <catalog> at PCL in NewsML-G2 v2.14:

- ❖ @url defines the location of the Catalog as a remote resource. (This must be the same as the URL used with the href attribute of a catalogRef in G2 Items that use this Catalog.)
- ❖ @authority uses a URI to define the company or organisation controlling this Catalog.
- ❖ @guid is a globally unique identifier, expressed as a URI, for this kind of Catalog as managed by a provider. (This must be the @guid of the Catalog Item, see below, that manages this Catalog) If present, the version attribute should also be used.
- ❖ @version is the version of the @guid as a non-negative integer; a version attribute must always be accompanied by a guid attribute.

14.6.3.2 <scheme> properties

Further information about schemes is expressed using the <scheme> child elements <name>, <definition>, and <note>. In NewsML-G2 v2.14 @roleuri was added to these child elements to allow the role of the element to be defined using a full URI instead of a QCode used by the existing @role attribute).

@roleuri simplifies processing by avoiding the situation where a QCode used in a Catalog relies on an alias defined in some other Catalog, making resolution difficult to impossible.

In NewsML-G2 v 2.17, the <schemeMeta> element was added to the Knowledge Item, enabling the metadata previously expressed in the <scheme> elements of a Catalog to be added directly to a Knowledge Item that conveys all the concepts in a Scheme (aka Controlled Vocabulary). See [13.3.5 Scheme Metadata](#)

14.6.4 Catalog Item

For providers who wish to use the same basic means for managing a Catalog as are available for news content, the Catalog Item is introduced to NewsML-G2 in version 2.14. The Catalog Item inherits the additions and changes to <catalog> and <scheme> described above.

The example below shows how a Scheme Authority (in this example, the IPTC) might distribute its catalog to subscribers.

The Catalog Item shares the generic identification and processing instruction attributes associated with all G2 items, for example:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<catalogItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20140517:catalog"
  version="22"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-GB">
```

The child properties of Catalog Item are restricted to a basic set of essentials required for Catalog management:

Definition	Name	Cardinality	Child properties
Catalog	catalogRef catalog	1..n	❖ As per all G2 Items
Hop History	hopHistory	0..1	❖ As per all G2 Items
Rights Information	rightsInfo	0..n	❖ As per all G2 Items
Item Metadata	itemMeta	1	❖ As per all G2 Items
Content Metadata	contentMeta	0..1	❖ contentCreated (0..1) ❖ contentModified (0..1) ❖ creator (0..1) ❖ contributor (0..1) ❖ altId (0..1) (power conformance only)
Catalog Container	catalogContainer	1	❖ catalog (1)

The catalogRef, rights information and itemMeta elements follow normal practice. Note that the Item Class is set to "cainature:catalog"; this uses the IPTC Catalog Item Nature NewsCodes, recommended Scheme Alias "cainature":

```
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<rightsInfo>
  <copyrightHolder uri="http://www.iptc.org">
```



```

    <name>IPTC</name>
  </copyrightHolder>
</rightsInfo>
<itemMeta>
  <itemClass qcode="cai:nature:catalog" />
  <provider qcode="nprov:IPTC">
    <name>International Press Telecommunications Council
    </name>
  </provider>
  <versionCreated>2014-05-17T12:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>

```

The Catalog information conveyed by the Item is wrapped in the <catalogContainer> element, which must contain one and only one <catalog>. The Catalog contains one or more <scheme> elements, as previously described:

```

<catalogContainer>
  <catalog xml:ns="http://iptc.org/std/nar/2006-10-01/"
    additionalInfo="http://www.iptc.org/goto?G2cataloginfo">
    <scheme alias="app" uri="http://cv.iptc.org/newscodes/application/">
      <definition xml:lang="en-GB">Indicates how the metadata
        value was applied.</definition>
      <name xml:lang="en-GB">Application of Metadata Values</name>
    </scheme>
  </catalog>
</catalogContainer>
</catalogItem>

```

LISTING 13 Complete Catalog Item

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<catalogItem xml:ns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20140517:catalog"
  version="22"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-GB">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <rightsInfo>
    <copyrightHolder uri="http://www.iptc.org">
      <name>IPTC</name>
    </copyrightHolder>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="cai:nature:catalog" />
    <provider qcode="nprov:IPTC">
      <name>International Press Telecommunications Council
      </name>
    </provider>
    <versionCreated>2014-05-17T12:00:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <catalogContainer>
    <catalog xml:ns="http://iptc.org/std/nar/2006-10-01/"
      additionalInfo="http://www.iptc.org/goto?G2cataloginfo">
      <scheme alias="app" uri="http://cv.iptc.org/newscodes/application/">
        <definition xml:lang="en-GB">Indicates how the metadata
          value was applied.</definition>
        <name xml:lang="en-GB">Application of Metadata Values</name>
      </scheme>
    </catalog>
  </catalogContainer>
</catalogItem>

```

14.7 Processing Catalogs and CVs

In practice, from a receiver's point of view, it makes no sense to look up the contents of CVs over the network every time a document is processed, since this would consume considerable computing and network resources and probably degrade performance. Also, as discussed, some providers might not make a scheme or its contents available at all.



The NewsML-G2 Specification requires that remote Catalogs – the file(s) that map Scheme Aliases to Scheme URIs – are retrieved by processors and the IPTC highly recommends that the Catalogs are cached at the receiver's site. They can be cached indefinitely because catalog URIs must remain unchanged over time. Whenever Schemes are created or deleted, an updated catalog must be provided under a new URI. This ensures that Items that pre-date the catalog changes can continue to be processed using the previous catalog.

14.7.1 Resolving Scheme Aliases

Some NewsML-G2 properties are important for the correct processing of an Item, for example the Item Class property tells a receiving application the type of content being conveyed by a NewsML-G2 Item: a processor may expect to apply some rule according to the value present in the <itemClass>, for example to route all pictures to the Picture Desk.

Other CVs may be important for correctly processing an Item, for example the presence of specific subject codes could cause an Item's content to be routed to certain staff or departments in a workflow.

The schemes used by <itemClass> property are mandatory, and the IPTC recommends that implementers use the scheme aliases appropriate to the Item Type, for example "ninat" for News item Nature or "cinat" for Concept Item Nature. Note that the use of these specific alias values is NOT mandatory; they could already being used by a provider as aliases for other CVs.

This illustrates the flexibility of the NewsML-G2 model: consistency of scheme aliases between different providers – or even by the same provider – cannot be guaranteed, and in NewsML-G2 they do not have to be guaranteed. For this reason it would be unwise for NewsML-G2 processor implementers to assume that a given scheme alias can be "hard coded" into their applications.

However, this flexibility does not mean that these "needed for processing" CVs must be accessed every time an Item is processed. This could be an unnecessary overhead and performance burden.

Processing rules such as those described above would be based on acting in response to expected values. In the case of the News Item Nature Scheme, these values include "text", "picture", "audio" etc. The problem is not in obtaining the contents of the CV in real time, but in verifying that it is the correct CV.

For example:

- ❖ A receiver knows that providers use the IPTC Media Topic NewsCodes for classifying news content by subject matter, and that the scheme URI for these NewsCodes is ["http://cv.iptc.org/newscodes/mediatopic/"](http://cv.iptc.org/newscodes/mediatopic/)
- ❖ The business requires that incoming content is routed to the appropriate department, according to the Media Topic NewsCodes found in the Items,
- ❖ A routing table is set up in the processor with a configurable rule "all items with a Media Topic NewsCode '04000000' to be routed to the Business News department".

How does the processor "know" that a <subject> property with a QCode containing "04000000" is an IPTC Media Topic NewsCode? The processor should not rely on the scheme alias "medtop"; it could be an alias to another CV, or the provider may use another alias:

```
<subject type="cpnat: abstract" qcode="sc: 04000000" />
```

By following the IPTC advice to retrieve all catalog information used by Items, and cache the information indefinitely, CV resolution can be performed in memory during processing.

In the example, the catalog used by the Item resolves the scheme alias "sc" contained in the QCode. The Item contains the line pointing to the catalog file:

```
<catalogRef href="http://www.example.org/std/catalog/catalog.example_10.xml" />
```

The processor should have retrieved and cached the contents of the file at this URL, and would have in memory the mapping of this alias to the Scheme URI:

```
<scheme alias="sc" uri="http://cv.iptc.org/newscodes/mediatopic/" />
```




...this verifies that the QCode value is from the Media Topic NewsCodes scheme. A rule “all items with a Media Topic NewsCode of ‘04000000’ to be routed to the Business News department” is satisfied and the NewsML-G2 Item processed appropriately.

14.7.2 Resolving Concept URIs

The IPTC recommends that Schemes SHOULD resolve to a Web resource, and that Scheme Authorities who disseminate news using NewsML-G2 should make their Schemes available as Knowledge Items.

14.7.2.1 Access Models

Making a CV available as a Web resource does not mean it must be accessible on the public Web; only that Web technology should be used to access it. The resource may be on the public Web, on a VPN, or internal network.

Providers may also wish to use Schemes to add value to content, using a subscription model. In this case, the contents of a Scheme may not be available to non-subscribers, but non-subscribers are still able to use the QCodes by mapping any QCode to a unique Concept URI that will be persistent over time.

14.7.2.2 Concept Resolution: Provider View

In following the IPTC recommendation that CVs should be accessible as a Web resource, providers may be concerned about the implications for providing sufficient access capacity and reliability guarantees. If receivers were to interrogate CVs each time they processed a NewsML-G2 document that could act like a Denial of Service attack.

The IPTC makes no recommendations about this issue other than to advise the use of industry-standard methods of mitigating these risks. Organisations hosting CVs could also define an acceptable use policy that places limits on the load that individual subscribers can place on the service.

14.7.2.3 Concept Resolution: Receiver View

As the IPTC recommends that CVs should be available as Web resources, it follows that the Scheme Authority may host its Schemes as Knowledge Items on a Web server. However, a Scheme Authority does not guarantee the availability and capacity of connections to its hosted Knowledge Items.

In addition, from the receiver's point of view, it could be unwise for business-critical news applications to rely on a third-party system beyond the receiver's control.

Processors are therefore recommended to retrieve and cache the contents of third-party Knowledge Items. Providers should advise their customers on the recommended frequency for refreshing the third-party cached Knowledge Items.

14.7.2.4 Handling updates to Knowledge Items using @modified

The receiver can get information about *which* concepts have been modified (when cached concepts are synchronized with those in the latest received Knowledge Item) using the modified attribute of <concept>

14.7.2.4.1 Use Case and Example

1. A news site is using a CV maintained by a third-party Scheme Authority, for example a CV maintained by the IPTC.
2. The site retrieves a Knowledge Item about the concepts in the CV from the third-party Scheme Authority's web server and stores them within its internal cache.
3. Sometime later the site wants to check the validity of the cache. It again downloads or receives a Knowledge Item from the third-party Scheme Authority, containing the relevant concepts which may have been updated in the meantime.
4. The site's NewsML-G2 processor checks the @modified timestamp (date-time) of each concept conveyed within the Knowledge Item against the modification timestamp of the corresponding cached concept. Any concepts within the Knowledge Item with a modification timestamp later than the corresponding cached

concept's modification timestamp are processed as updates to the cache. (Note: this assumes that the Scheme Authority always flags Concepts conveyed within a Knowledge Item with a modification timestamp, see below.)

14.7.2.4.2 Updates Processing Notes

In the above use case, it was assumed that the Scheme Authority always flags Concepts with a modification timestamp. In cases where modification timestamps are missing from some or all of the concepts, either in the KI or in the cache, a receiver can be less certain about whether or not a concept has been modified. The following matrix outlines the IPTC recommendations for processing updates for each individual concept in the KI:

Concept in local cache	Concept just received in KI	Processor should:
No modification timestamp	No modification timestamp	Update cache from KI
No modification timestamp	Has modification timestamp	Update cache from KI, now the concept in the cache has a modification timestamp!
Has modification timestamp	No modification timestamp	Update cache from KI, now the concept in the cache has lost its modification timestamp!
Has modification timestamp	Has modification timestamp	Compare modification timestamp: <ul style="list-style-type: none">i. If the modification timestamp from the KI is later than the one from the cache: update the concept in the cache.ii. if the modification timestamp from the KI is earlier than the one from the cache: do not update the concept in the cache.

The code snippet below shows how the Scheme Authority would inform receivers that the concepts in a Knowledge Item have been updated using the @modified attribute value informing receivers of the timestamp (date-time) of the change.

```
<?xml version="1.0" encoding="UTF-8"?>
<knowledgeItem ... >
  ...
  <itemMeta>
    ...
  </itemMeta>
  <contentMeta>
    ...
  </contentMeta>
  <conceptSet>
    <concept modified="2010-01-28T13:00:00Z">
      <conceptId qcode="access:easy" />
      ...
    </concept>
    <concept modified="2010-01-28T13:00:00Z">
      <conceptId qcode="access:difficult" />
      ...
    </concept>
    <concept modified="2009-11-23T13:00:00Z">
      <conceptId qcode="access:restricted" />
      ...
    </concept>
  </conceptSet>
</knowledgeItem>
```

14.7.2.4.3 Notifying receivers of changes to Knowledge Items

This issue is not necessarily specific to NewsML-G2 news exchange: Most news providers have CVs that pre-date NewsML-G2, for example those CVs typically used with IPTC 7901. Channels and conventions



for advising customers of changes to CVs will already exist. Generally, providers notify customers in advance about changes to CVs, especially if it is likely that a CV is used for content processing.

The IPTC hosts and maintains a large number of CVs and provides an RSS feed that notifies of changes to the IPTC Schemes. Details at www.iptc.org.

14.8 Private versions and extensions of CVs

News providers are encouraged to use pre-existing or well-known CVs, such as those maintained by the IPTC, where possible to promote interoperability and standardisation of the exchange of news. Sometimes a provider will use a CV that is maintained by some other Scheme Authority (e.g. the IPTC), but may need to add its own information. The following are typical potential business cases:

- ❖ **Case 1:** A national news agency wishes to use all of the codes in a CV that is maintained by the IPTC, without alteration, except to provide local language versions of names and definitions.
- ❖ **Case 2:** An organisation receives news objects from information partners and uses a CV that defines the stages in a shared workflow. The CV is maintained by a third-party organisation (the Scheme Authority), but the receiver needs to add further workflow stages for its own internal purposes.

14.8.1 Use <schemeSameAs> to provide a local language version of a well-known CV

Some useful and well-known CVs do not have the values of properties such as name and definition in the local language of a news provider. For example, although some IPTC NewsCodes contain concept information in several languages, the IPTC does not have the resources to provide concept details in every language being used by news providers.

Yet some NewsCode schemes are recommended or mandatory. How can a provider use a local language version of these NewsCodes AND conform to the NewsML-G2 Specification?

Using the <sameAsScheme> property, a news provider can create its own CV, containing all of the NewsCodes it wishes to use with local language names and definitions, while making it clear to receivers that these codes identify the same concepts as the original IPTC Scheme.

For example, the IPTC Item Relation NewsCode “irel:seeAlso” resolves via the Catalog to a Concept URI “http://cv.iptc.org/newscodes/itemrelation/seeAlso” which hosts the following information about the seeAlso concept:

```
<concept>
  <conceptId created="2008-01-29T00:00:00+00:00" qcode="irel:seeAlso" />
  <type qcode="cpnat:abstract" />
  <name xml:lang="en-GB">See also</name>
  <definition xml:lang="en-GB">
    To fully understand the content of this item see also the content
    of the related item.
  </definition>
  ...
</concept>
```

A provider wishes to provide this same information in the Czech language. As a first step, it creates a new Controlled Vocabulary containing the required concepts from the original scheme with translated name and definition, for example:

```
<concept>
  <conceptId created="2010-01-29T00:00:00+00:00" qcode="itemrel:seeAlso" />
  <type qcode="cpnat:abstract" />
  <name xml:lang="cs">Viz také</name>
  <definition xml:lang="cs">
    Chcete-li plně pochopit obsah této položky viz též obsah
    související položky.
  </definition>
  ...
</concept>
```

The provider then creates a Catalog file, or a new version of an existing Catalog file, containing a Scheme Alias and Scheme URI for the new CV, thus:

```
<catalog>
  <scheme alias="itemrel" uri="http://cv.example.org/codes/itemrelation/">
```

```
<sameAsScheme>http://cv.iptc.org/newscodes/itemrelation/</sameAsScheme>
</scheme>
</catalog>
```

This asserts that the ALL of the codes in the private scheme identified by the Scheme URI are semantically the “same as” the corresponding codes in the original scheme indicated by the <sameAsScheme> child element of <scheme>.

Note that in this example the provider MUST give the new CV a Scheme Alias – here it is “itemrel” – that is different to the recommended Scheme Alias “irel” of the IPTC Scheme. This is because some IPTC schemes are mandatory, so a reference to the IPTC catalog would always be present in the Item. When there is no reference to an original scheme, there is no need to use a different Scheme Alias for the private scheme.

Finally, the provider adds a reference to the new Catalog file to NewsML-G2 Items that it publishes.

By using the <sameAsScheme> element to the Catalog, the provider is able apply a Same As relationship at the level of a **set** of Concepts. So for example, this code:

```
<link
  rel="itemrel:seeAlso"
  contentType="image/jpeg"
  residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
/>
```

can be efficiently resolved: a processor does not have to search for a Same As relationship at the individual concept level but can map this relationship directly from the @rel value’s scheme (alias “itemrel”) to the scheme identified by the <sameAsScheme> property.

14.8.1.1 Rules for <sameAsScheme>

The semantics of <sameAsScheme> are:

“all of the concepts in the scheme identified by the private scheme/@uri have a ‘same as’ relationship to concepts with the same code in the original scheme identified by the URI in the <sameAsScheme> element.”

So in the example:

```
<scheme alias="itemrel" uri="http://cv.example.org/codes/itemrelation/">
  <sameAsScheme>http://cv.iptc.org/newscodes/itemrelation/</sameAsScheme>
</scheme>
```

In practice, this means:

- ❖ The Scheme identified by scheme/@uri (the *provider’s* private scheme) must NOT use a code that does not exist in ALL of the *original* Schemes identified by the <sameAsScheme> elements. In other words, a provider cannot add new concepts to its private Scheme that have the effect of extending the set of concepts of the original Scheme(s).
- ❖ Some codes and concepts of the original Scheme MAY not exist in the provider’s private Scheme. This could happen if for example the original Scheme has new terms added which the provider has not yet included in the private Scheme.
- ❖ Each concept identified by a code in the provider’s private Scheme MUST be semantically equivalent to its corresponding concept in the original Scheme(s), and MUST be identified by the same code as in the original Scheme(s).

The <sameAsScheme> property was introduced to solve some issues that news providers had encountered, such as adding translations of free-text properties (name, definition, note etc), which are not available within the original scheme, or adding additional information, e.g. usage notes.

14.8.2 Adding further concepts to a well-known CV

A news provider needs to add concepts to a workflow role CV which is shared with information partners, but is maintained by some other organisation (the Scheme Authority).

The news provider would have two courses of action:

- a) Ask the Scheme Authority to add the new concepts to the CV. For example, IPTC members are entitled to request the addition and/or retirement of terms in IPTC Schemes with the agreement of other members.
- b) Create a new scheme that complements the original scheme, but uses properties such as `<broadener>` and `<sameAs>` to link the concepts in the new scheme to concepts in the original scheme. A concept is the Same As another concept if their semantics are the same, but it MAY contain more details, such as a translation in another language. A concept with a Broadener relationship to another concept is a new concept with semantics narrower than those of the broader concept.

14.8.2.1 Example using a new scheme

Using the shared workflow role example, the original scheme contains three concepts for defined roles in a workflow:

```
<conceptSet>
  <concept>
    <conceptId qcode="wflow: draft" />
    ...
  </concept>
  <concept>
    <conceptId qcode="wflow: review" />
    ...
  </concept>
  <concept>
    <conceptId qcode="wflow: release" />
    ...
  </concept>
</conceptSet>
```

Properties that use this scheme are resolved through the `<catalog>` in the Item in which they appear, e.g. the Item contains the following property in `<itemMeta>`:

```
<role qcode="wflow: release"
```

and the catalog statement:

```
<catalog>
  <scheme alias="wflow" uri="http://cv.example.org/schemes/wflowes/" />
</catalog>
```

The receiver needs to add an intermediate role in the workflow, representing a "final review" stage. Thus the private scheme in Case 2 is extending the original scheme. Unlike Case 1, where the codes in the private scheme were identical to codes contained in the original scheme, In Case 2, the concepts in the private scheme use a `<sameAs>` property at the level of each code in the new private scheme.

```
<conceptSet>
  <concept>
    <conceptId qcode="iwf: draft" />
    <sameAs qcode="wflow: draft" />
    ...
  </concept>
  <concept>
    <conceptId qcode="iwf: review" />
    <sameAs qcode="wflow: review" />
    ...
  </concept>
  <concept>
    <conceptId qcode="iwf: final review" />
    ...
  </concept>
  <concept>
    <conceptId qcode="iwf: release" />
    <sameAs qcode="wflow: release" />
    ...
  </concept>
</conceptSet>
```

G2 Items that use this scheme must use a `<catalog>` statement to enable the processor to resolve both the private "iwf" scheme alias and the original "wflow" scheme alias:

```
<catalog>
```

```
<scheme alias="iwf" uri="http://support.myorg.com/cv/workflow/" />
<scheme alias="wflow" uri="http://cv.example.org/schemes/wfroles/" />
</catalog>
```



A Knowledge Item containing <sameAs>, <broader>, or <narrower> properties like the above must also contain a <catalog> allowing the QCodes to be resolved, in this case:

```
<catalog>
<scheme alias="wflow" uri="http://cv.example.org/schemes/wfroles/" />
</catalog>
```

14.9 Best Practice in QCode exchange

NewsML-G2 specifies that concepts must be identified by a full URI conforming to RFC 3986. The IPTC also recommends that a URI identifying a scheme and concept should resolve to a resource providing information about the scheme or the concept and which is either human or machine readable. In other words, a Concept URI should be a URL.

The unreserved characters that are permitted in a URL are:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o p q r s t u v w x y z
0 1 2 3 4 5 6 7 8 9 - _ . ~

the reserved characters are:

! * ' () ; : @ & = + \$, / ? % # []

In order to promote unambiguous processing of QCodes, the IPTC defines that it is the responsibility of providers to encode Concept URIs as they are intended to be valid for a system which resolves them and not to rely on end-users applying per-cent encoding rules for URI processing. Therefore:

- ❖ Providers should decide whether or not to per-cent encode reserved characters used in QCodes distributed to customers in G2 documents.
- ❖ Receivers should not perform per-cent encoding/decoding, when resolving QCodes according to the rules outlined below.

The reason for the recommendation is illustrated by the following example of a provider distributing a document containing:

```
<subject qcode="fc: 3#FTSE" />
```

The catalog entry for a parent Scheme is (say):

```
<scheme alias="fc" uri="http://cv.example.org/schemes/fc/" />
```

The receiver could transform this QCode literally to this URI:

```
http://cv.example.org/schemes/fc/3#FTSE
```

Or the receiver could percent-encode the # (%23), yielding:

```
http://cv.example.org/schemes/fc/3%23FTSE
```

The resources identified by these two URIs are both valid by RFC 3986 but they are different!

Following the IPTC definition, the provider must include the per-cent encoding into the QCode of the first example if this is required by the system resolving the URIs:

```
<subject qcode="fc: 3%23FTSE" />
```

14.9.1 Non-ASCII characters

The encoding described above assumes that the character(s) to be per-cent encoded are from the US-ASCII character set (consisting of 94 printable characters plus the space).

If a code contains non-ASCII characters, for example accented characters, the Unicode encoding UTF-8 must be used, in line with normal practice.

For example, the UTF-8 encoding of the “å” character is a two-byte value of C3_{hex} A5_{hex} which would be percent-coded as “%C3%A5”.

14.9.2 White Space in Codes

Codes in controlled vocabularies which have been created with no regards to the NewsML-G2 specifications may contain spaces. As the NewsML-G2 Specification does not allow white space characters in Codes, this section recommends a workaround.

Whitespace characters in Codes - in practice, only spaces (20_{hex}) - are replaced by a sequence of one or more unreserved characters that is reused for this purpose according to the practices of the provider; it is recommended that such a sequence is not part of the any of the codes used by the provider.

For example, if a code contains a space, the space character might be replaced by ~. Receivers would be informed to translate this string back to a space character in order to match the QCode against a list of codes that contain spaces.

14.10 Syntactic Processing of QCodes

This section provides a summary of the processing model. Please also see the NewsML-G2 Specification for a full technical description. (If not at hand, this can be downloaded by visiting www.newsml-g2.org and navigating to Specification resources.)

14.10.1 Creating QCodes from Scheme Aliases and Codes

14.10.1.1 Scheme Aliases

These do not have to be encoded as they will never be part of the full Concept URI. A Scheme Alias may contain any character except a colon (3A_{hex}) or white space characters (20_{hex} or 09_{hex} or 0D_{hex} or 0A_{hex})

14.10.2 Processing Received Codes

To resolve a QCode received in an Item to a Concept URI, use the following steps:

1. Apply any XML decoding to the string (this should be performed by your XML processor)
2. Retrieve the QCode value from the document
example: fôô:bår
3. Identify the first colon starting from the left; the string on left of the colon is the scheme alias, the string on the right of the colon is the code. If there is no colon, the QCode is invalid. In the example, therefore:
Scheme Alias = fôô
Code Value = bår
4. Check whether the alias is defined in a catalog. If not, the QCode is invalid.
example: <scheme alias="fôô" uri="http://cv.example.org/cv/somecodes/" />
5. Append the Code Value to the Scheme URI to make the full Concept URI:
example: http://cv.example.org/cv/somecodes/bår
6. It is highly recommended to use only full Concept URIs to compare identifiers of concepts.

14.11 A generic way to express concept identifiers as URIs

When NewsML-G2 was originally designed, communication capacity constraints drove implementers to find a compact way to express the typically long globally unique identifier (URI) of a concept; this led to the creation of the compact QCode notation. In NewsML-G2 version 2.11, @uri was introduced for properties having a concept identifier value type. This was restricted to properties that used @qcode or @literal to hold the identifier, but additionally allowed providers to express the full URI if required. Example:

```
<subject uri="http://example.com/people?id=12345&group=223" />
```

To make the same flexibility available to properties using the QCodeType or QCodeList type of attribute, the new URI sibling attributes (IRIType) have now been added to all properties that take these types of



attribute (QCodeType or QCodeListType) by adding "uri" to the base name of the attribute. For example @role gets a @roleuri sibling, such that:

```
<group id="G1" role="group:main">
```

may now be expressed as:

```
<group id="G1" roleuri="http://cv.example.org/grouprole/main">
```

Implementers should NOT use both QCode and URI type attributes for the same property; however, if they are used together, the QCode type attribute takes precedence. For a few properties @qcode is mandatory; in such cases both a QCode value AND a full URI value MAY be used.

The following extract shows how the <contentSet> of the code listing from **Quick Start: Pictures and Graphics** would use the uri sibling properties:

```
<contentSet>
  <remoteContent renditionuri="http://cv.iptc.org/newscodes/rendition/highRes"
    href="/GYI0062134533.jpg"
    version="1"
    size="346071"
    contentType="image/jpeg"
    width="1500"
    height="1001"
    colourspaceuri="http://cv.iptc.org/newscodes/colourspace/AdobeRGB"
    orientation="1"
    layoutorientationuri="http://cv.iptc.org/newscodes/layoutorientation/horizontal">
    <altId type="gyiid:masterID">105864332</altId>
  ...
</contentSet>
```

This code is listed in full in the file *LISTING 3A Photo in NewsML-G2 (URI sibling attributes).xml*, which is part of the set of code examples in the Guidelines download package.

See **14.13 Full list of the QCode type attributes and their URI siblings** (below)

14.12 Literal Identifiers

The NewsML-G2 Standard recognises that it is not always possible to use a QCode or URI as an identifier, therefore the G2 Flexible Property type allows a @literal identifier, or no identifier at all. For example, a CV may be understood by both receivers and providers, but the mapping of identifiers to concepts is managed and communicated outside NewsML-G2. Many long-established CVs such as these pre-date NewsML-G2. In other circumstances, the identifier may add no value, because only some basic property, such as <name> needs to be conveyed.

A @literal is an identifier which is intended to be processed by software; it is not intended to be a natural-language label. If the name of a concept identified by @literal is intended for display, the IPTC recommends that providers SHOULD add the <name> child element for inter-operability and language-independent processing. If no human-readable property is available, receivers MAY use the @literal value for display purposes.

14.12.1 Use of @literal in a NewsML-G2 Item

Literals may be used as in the following cases:

1. As an identifier for linking with an assert element inside a NewsML-G2 document. In this case the literal value could be a random one. If a literal value is used with an assert element then all instances of that literal value in that item must identify the same concept.
2. When a code from a vocabulary which is known to the provider and the recipient is used without a reference to the vocabulary. The details of the vocabulary are, in this case, communicated outside NewsML-G2. Such a contract could express that a specific vocabulary of literals is used with a specific property.
3. When importing metadata which may contain codes which have not yet been checked to be from an identified vocabulary: the code values are represented as literals until the vocabulary is identified; thereafter, a controlled identifier can be used.

The following rules govern the use of @literal:

- ❖ A @literal value can only be used to identify a concept within the local scope of an Item.

- ❖ The use of @literal and a controlled identifier (either @qcode, @uri or both) is mutually exclusive.
- ❖ There can be no guarantee that all instances of a @literal value used in an Item identify the same concept. However, when @literal is used with an assert element, providers MUST ensure all instances of that literal value in the Item identify the same concept. If the provider uses the same literal value for different concepts, an assert element using this literal value MUST NOT be used, as the concept is indeterminate.

14.12.2 Properties with no identifier

It is permissible for a NewsML-G2 element with a Flexible Property type to have no concept identifier:

```
<provider>
  <name>Getty Images North America</name>
</provider>
```

As a special case, when a <bag> child element is used with a property to create a composite concept, NO concept identifier may be used with the parent property (neither @qcode, nor @uri nor @literal). The composite concept of the <bag> uses multiple existing concepts which MUST be identified by @qcode attributes of the <bit> elements to create a new one:

```
<subject>
  <name>Bread</name>
  <bag>
    <bit qcode="ingredient: flour"/>
    <bit qcode="ingredient: water"/>
    <bit qcode="ingredient: yeast"/>
  </bag>
</subject>
```

14.13 Full list of the QCode type attributes and their URI siblings

Attribute Name	URI Sibling Attribute	Notes
aspect	aspecturi	
audiochannels	audiochannelsuri	
audiocodec	audiocodecuri	
colourindicator	colourindicatoruri	
colourspace	colourspaceuri	
creator	creatoruri	
durationunit	durationunituri	
encoding	encodinguri	
environment	environmenturi	
format	formaturi	
hashtype	hashtypeuri	(required in <hash>)
heightunit	heightunituri	
how	howuri	
idformat	idformaturi	
interactiontype	interactiontypeuri	(required in <userinteraction>)
jobtitle	jobtitleuri	
layoutorientation	layoutorientationuri	
market	marketuri	
marketlabelsrc	marketlabelsrcuri	



Attribute Name	URI Sibling Attribute	Notes
mode	modeuri	
part	parturi	
pubconstraint	pubconstrainturi	
radunit	radunituri	(required in <circle>)
ratertype	ratertypeuri	
ratingtype	ratingtypeuri	
rel	reluri	(required in all <...ExtProperty>)
rendition	renditionuri	
renditionref	renditionrefuri	
representation	representationuri	
role	roleuri	(required in <group>)
scaleunit	scaleunituri	(required in <rating>)
scope	scopeuri	
severity	severityuri	
symbolsrc	symbolsrcuri	
target	targeturi	
tech	techuri	
timeunit	timeunituri	(required in <timedelim>)
type	typeuri	
valcalctype	valcalctypeuri	
valueunit	valueunituri	
videocodec	videocodecuri	
videodefinition	videodefinitionuri	
videoscaling	videoscalinguri	
why	whyuri	
widthunit	widthunituri	



15 EventsML-G2

15.1 A standard for exchanging news event information

The sharing of event-related information and planning of news coverage is a core activity of news organisations, without which they cannot function effectively.

News agencies need to keep their customers *informed* of upcoming events and planned coverage. News organisations publishing on paper and digital media need to *plan and co-ordinate* their operations in order to make optimum use of their available resources and ensure their target audiences will be properly served.

Historically, this was a paper-based exercise, with news desks maintaining a Day Book, or Diary, and circulating information to colleagues and partners using written memoranda, sometimes referred to as the Schedule, or Budget.

Many organisations have moved, or are moving, to electronic scheduling applications. With software developers and vendors working independently on these applications, there is a risk that incompatibility will inhibit the exchange of information and reduce efficiency.

Consequently, there has been a drive among IPTC members to formalise a standard for exchanging this information in a machine-readable events format using XML, allowing it to be processed using standard tools and enabling compatibility with other XML-based applications and popular calendaring applications such as Microsoft Outlook.

As an Events Calendar and Scheduling model, EventsML-G2 is focussed on the needs of a professional news industry workflow. There is the need to express the basic news agenda of “what, when, where and who” in a machine-readable form that can be processed by calendar and planning tools.

The related NewsML-G2 Planning Item can be combined with Event information so that news organisations can plan their response to news events, such as job assignments, content planning and content fulfilment, sharing this if required with partners in a workflow. (See [Editorial Planning – the Planning Item](#))

EventsML-G2 is used to send and receive all, or part of, the information about:

- ❖ a specific news event.
- ❖ a range of news events filtered according to some criteria – an event listing.
- ❖ updates to news events.
- ❖ people, organisations, objects and other concepts linked to news events.

15.1.1 Business Advantages

EventsML-G2 shares a common structure with NewsML-G2⁶. Thus an event managed within the G2 Family can serve as the “glue” that can bind together all of the content related to a news story.

For example, a news organisation learns of the imminent merger of two important companies. This story can be pre-planned using EventsML-G2 and assigned a unique Event ID by the event planning workflow, in the form of a QCode. When content related to this event is created (text, pictures, graphics, audio, video and packages), all of the separately-managed content to be associated using the QCode as a reference.

The result is that when users view content about this story, they can be provided with navigation to any other related content, or they can search for the related content.

EventsML-G2 is the result of detailed collaborative work by IPTC experts operating in diverse markets throughout North America, Europe and Asia Pacific. The EventsML-G2 Working Group is highly experienced in the planning of news operations and the issues involved.

Adopters therefore have access to an “off the shelf” data model built on the specific needs of the news industry that can nevertheless be extended by individual organisations where necessary to add specialised features. EventsML-G2 will also evolve as new requirements become known, and the IPTC endeavours to maintain compatibility with previous versions if possible, giving users a straightforward upgrade path.

⁶ From v2.9, the two standards have in fact been merged into one.

EventsML-G2 complies with the Resource Description Framework (RDF) promulgated by the W3C, which is a basic building block of the Semantic Web, and aligns with the iCalendar⁷ specification that is supported by popular calendar and scheduling applications such as Microsoft Outlook, Lotus Notes and Apple iCal.

There is considerable scope for using events planning to improve the efficiency and quality of news production, since an estimated 50-80% of news provision is of events that are known about in advance. When a pre-planned event is accessible from an editorial system, metadata from the event may be inherited by the news content associated with that event; this makes the handling of the news faster and more consistent. There is also an improvement in quality, since the appropriateness and accuracy of metadata may be checked at the planning stage, rather than under the pressure of a deadline.

The advantages of using a common standard to promote the efficient exchange of information are well understood. Using EventsML-G2, providers can develop planning and scheduling products with greater confidence that the information can be consumed by their customers; recipients can cut development costs and time to market for the savings and services that flow from an efficient resource planning system that aligns with their operational model.

15.1.2 Structure and compatibility

EventsML-G2 is part of the G2 family derived from the News Architecture (NAR) data model. After version 1.7, it has been merged with NewsML-G2 – they have a common XML Schema, with NewsML-G2 being extended to accommodate the event-specific features of EventsML-G2. The version numbers are in step as of version 2.9 and beyond. The change is mainly a simplification of the number of Schemas that implementers need to adopt; the following applies to all versions of EventsML-G2:

- ❖ Events use NewsML-G2 identification and versioning properties.
- ❖ The <itemMeta> block holds management information about the G2 object.
- ❖ The <contentMeta> block holds common administrative metadata about the event, or events, conveyed by the NewsML-G2 object.



As there is no longer a separate EventsML-G2 standard, news event management is now also an integral part of NewsML-G2.

Please read the Chapter on NewsML-G2 Basics before reading this Chapter.

There are two methods for expressing events using G2, each suited to a particular type of information, or application, as shown in **Figure 14**:

- ❖ Persistent event information that will be referenced by other events and by other G2 Items is instantiated as Concept Items, or as members of a set of event Concepts in a Knowledge Item. Events expressed as a Concept with a Concept ID may be persistently and unambiguously referenced by other Items over time.
- ❖ Transient event information that is “standalone” and volatile can be conveyed in NewsML-G2 using an <events> structure inside a News Item. Events expressed as an <event> in a News Item, have no Concept ID and can never be referenced from other Items.

Managed (persistent) events would be appropriate when a provider makes each of the announced events uniquely identifiable. The event information can then be stored by the receivers and any updates to events can be managed. This model also enables content to be linked to events using the unique event identifier, and this enables linking and navigation of content related to news stories.

A “standalone” implementation would be used when an organisation periodically announces to its partners and customers lists of forthcoming events. These may be for information only and not managed by the provider. So, for example, a daily list may repeat items that appeared in a weekly list, but there is no link between them, and nothing to indicate when an event has been updated.

The difference in NewsML-G2 terms between the two types of implementation are illustrated in the following diagram:

⁷ A mapping of iCalendar properties to G2 properties can be found in the [G2 Specification document](#) on the IPTC web site. See also [IETF iCalendar Specification RFC 5545](#)

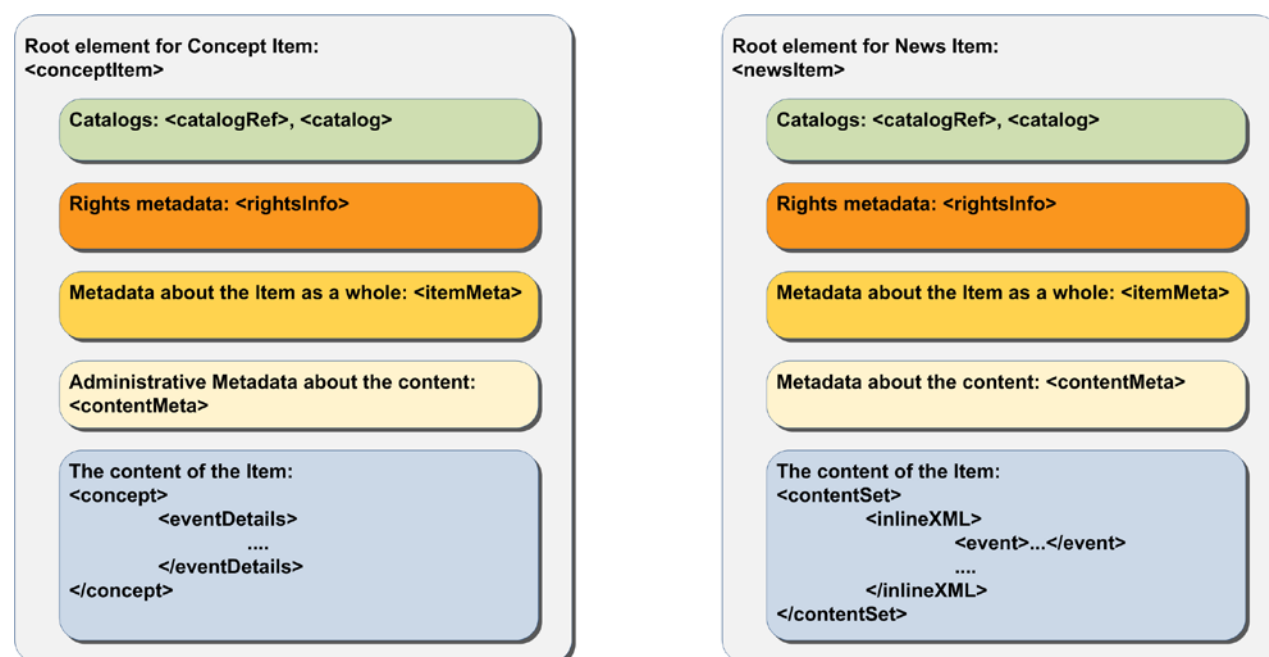


Figure 14: Persistent Event as a Concept using `<eventDetails>` and (right) transient Events in a News Item

15.2 Event Information – What, Where, When and Who

In a news context, events are newsworthy happenings that may result in the creation of journalistic content. Since news involves people, organisations and places, G2 has a flexible set of properties that can convey details of these concepts. There is also a fully-featured date-time structure to express event occurrences, which conforms to the iCalendar specification.

The following example shows a news event expressed as a Concept and conveyed within a Concept Item.

The top level element of the Concept Item is `<conceptItem>`. The document must be uniquely identified using a GUID. By this means, event information re-sent using the same GUID and an incremented version number, allows the receiver to manage, update or replace the conveyed concept (event) information.



@guid and @version uniquely identify the Concept Item, for the purpose of managing and updating the event information. Items that reference the event itself MUST use the Concept ID. This is because the Concept ID uniquely references a persistent Web resource, whereas the GUID only identifies a document that may or may not persist.

To enable concepts to be identified by a Concept ID QCode, a reference to the provider's catalog (or a catalog statement containing the scheme URI) MUST be included:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20090122:qqwpi ruuew4711"
  version="2"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-standards_25.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
```

In the mandatory `<itemMeta>` wrapper we use the IPTC "Nature of Concept Item" NewsCode to express the type of concept item being conveyed. (This is complementary to the "Nature of News Item" NewsCode used with a News Item.) There are currently two values: "concept" and "scheme". (Scheme is used for Knowledge Items.)

```
<itemMeta>
  <itemClass qcode="ci nat: concept" />
  <provider qcode="nprov: IPTC" />
  <versionCreated>2012-10-18T13:05:00Z
</versionCreated>
  <pubStatus qcode="stat: usable" />
```



```
</itemMeta>
```

The Content Metadata for a Concept Item may contain only Administrative Metadata:

```
<contentMeta>
<urgency>5</urgency>
<contentCreated>2012-10-16T12:15:00Z</contentCreated>
<contentModified>2012-10-16T13:35:00Z</contentModified>
</contentMeta>
```

15.2.1 What is the Event?

In order to convey event information, we first need to describe “what” the event is. In G2, events **MUST** have at least one Event Name, a natural language name for the event. They **MAY** additionally have one or more natural language Definitions, properties that describe characteristics of the event, and Notes.

This Event content is contained in the <concept> wrapper. A Concept Item has one, and only one, <concept> wrapper. Each Concept must carry a <conceptId>, a QCode value that will provide a way for other Items to reference the Concept. The type of concept being conveyed is expressed using the <type> element with a QCode to tell the receiver that it contains Event information, using the IPTC “Nature of Concept” NewsCode. (The applicable value for an Event Concept is “event”) The Scheme URI is <http://cv.iptc.org/newscodes/cpnature/> and recommended Scheme Alias is “cpnat”.

The event information is placed directly inside an <eventDetails> wrapper:

```
<concept>
  <conceptId created="2012-10-16T12:15:00Z" qcode="event:1234567" />
  <type qcode="cpnat:event" />
  <name>IPTC Autumn Meeting 2012</name>
  <eventDetails>
    ...
  </eventDetails>
</concept>
```

The generic properties of Events are similar to those of Concepts, covered in [Concepts and Concept Items](#). In this framework, we can expand the “what” information of an Event by indicating one or more relationships to other events, using the properties of Broader, Narrower, Related and SameAs. (See [15.4.2 Event Relationships](#))

15.2.2 When does the Event take place?

The “when” of an event uses the Dates wrapper to express the dates and times of events: the start, end, duration and recurrence.

```
<dates>
  <start>2012-10-22T09:00:00-04:00</start>
  <duration>P3D</duration>
</dates>
```

Although start and end times may be specified precisely, in the real world of news, the timing of events is often imprecise. In the early stages of planning an event, the day, month or even just the year of occurrence may be the only information available. Providers also need to be able to indicate a range of dates and/or times of an event, with a “best guess” at the likely date-time. (See [15.4.3.1 Dates and times](#) for details)

15.2.3 Where does the Event take place?

The “where” of an event is expressed using Location, a rich structure containing detailed information of the event venue (or venues); including GPS coordinates, seating capacity, travel routes etc.

```
<location>
  <name>Hotel Mercure Arbat Moscow, Smolenskaya PL 6, 121099 Moscow</name>
  <related rel="frel:venue" qcode="ventyp:hotel" />
  <POIDetails>
    <position latitude="55.74906" longitude="37.583109" />
    <contactInfo>
      <web>http://www.accorhotels.com/gb/7454-mercure-arbat-moscow/</web>
    </contactInfo>
  </POIDetails>
</location>
```



Note that if in-depth location details are given in the Event structure, this is a “one-time” use of the information. This structure might be better used as part of a controlled vocabulary of locations, in which case the structures may be copied from a referenced concept containing the location information.

See **15.4.3.2 Further Properties of Event Details** for details of Location and a comprehensive list of other Event properties.

15.2.4 Who is involved?

Details of “who” will be present at an event are given using the Participant property. This can be expressed simply using a QCode, URI or literal value, supplemented by a human-readable Name property, or optionally using additional properties describing the participants, their roles at the event, and a wealth of related information.

```
<participant qcode="iptcofficer:msteidl">
  <name>Michael Steidl</name>
  <personDetails>
    <contactInfo>
      <email>mdirector@iptc.org</email>
    </contactInfo>
  </personDetails>
</participant>
```

The event organisers are also an important part of the “who” of an event. A set of Organisation and Contact Information properties can give precise details of the people and organisations responsible for an event, and how to contact them. See **15.4.3.2 Further Properties of Event Details** for details

15.2.5 Completed Listing for Example Event

LISTING 14 Event sent as a Concept Item

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20090122:qqwpi ruuew4711"
  version="2"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-standards_25.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider qcode="nprov:IPTC" />
    <versionCreated>2012-10-18T13:05:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2012-10-16T12:15:00Z</contentCreated>
    <contentModified>2012-10-16T13:35:00Z</contentModified>
  </contentMeta>
  <concept>
    <conceptId created="2012-10-16T12:15:00Z" qcode="event:1234567" />
    <type qcode="cpnat:event" />
    <name>IPTC Autumn Meeting 2012</name>
    <eventDetails>
      <dates>
        <start>2012-10-22T09:00:00-04:00</start>
        <duration>P3D</duration>
      </dates>
      <location>
        <name>Hotel Mercure Arbat Moscow, Smolenskaya PL 6, 121099 Moscow</name>
        <related rel="frel:venue" qcode="ventyp:hotel" />
        <POIDetails>
          <position latitude="55.74906" longitude="37.583109" />
          <contactInfo>
            <web>http://www.accorhotels.com/gb/7454-mercure-arbat-moscow/</web>
          </contactInfo>
        </POIDetails>
      </location>
    </eventDetails>
  </concept>
</conceptItem>
```




```
<name>Michael Steidl</name>
<personDetails>
  <contactInfo>
    <email>mdirector@iptc.org</email>
  </contactInfo>
</personDetails>
</participant>
</eventDetails>
</concept>
</conceptItem>
```

15.3 Event Coverage

The G2 Planning Item is used to inform customers about content they can expect to receive and if necessary the disposition of staff and resources. (see [Editorial Planning – the Planning Item](#)).⁸

The link between an Event and the optional Planning Item is created using the Event ID. The Planning Item conveys information about the content that is planned in response to the Event, including the types of content and quantities (e.g. expected number of pictures). The Planning Item also has a Delivery structure which enables the tracking and fulfilment of content related to an Event.

15.4 Event Properties in more detail

The examples below show the basic properties of events, starting with the Name and Definition of the event, with other descriptive details, moving on to show how relationships, location, dates and other details may be expressed.

15.4.1 Event Description

In the code sample that follows, we will begin to create an event using four properties:

- ❖ Event Name is an internationalised string giving a natural language name of the event. More than one may be used, for example if the Name is expressed in multiple languages.
- ❖ Definitions, two differentiated by @role (“short” and “long”) will be created using the Block element template (which allows some mark-up).
- ❖ Notes – also Block type elements – give some additional natural language information, which is not naturally part of the event definition, again using @role if required.
- ❖ Using Related, the nature of the Event can be refined and other properties of an Event can be expressed.

Property	Type	Notes/Example
<name>	Intl String	<pre><name xml:lang="en"> Bank of England Monetary Policy Committee </name></pre>
<definition>	Block	<pre><definition xml:lang="en" role="drole: short"> Monthly meeting of the Bank of England committee that decides on bank lending rates for the UK. </definition> <definition xml:lang="en" role="drole: long"> The Bank of England Monetary Policy Committee meets each month to decide on the minimum rate of interest that will be charged for inter-bank lending in the UK financial markets.
 The committee will discuss the prospects for economic activity in the UK and overseas markets, and make its decision on interest rates based on forecasts for</pre>

⁸ In EventsML-G2 prior to version 1.6, the <eventDetails> wrapper used a <newsCoverage> child element to convey editorial planning information. This use of <newsCoverage> is now DEPRECATED and the same News Coverage structure is now available in the Planning Item. One of the many advantages of this separation is that it allows frequent updates to Planning to be sent without affecting the Event Concept to which it refers.

		inflation, amongst other factors. </definition>
<related>	Flexible	<p>The nature of the event, and other information. The property is repeatable and may have a @rel attribute to indicate the relationship to the characteristic. Relationships can be expressed using:</p> <ul style="list-style-type: none"> ❖ a QCode referencing a controlled vocabulary of (for example) event types, such as “meeting”, “parliamentary session”, “music concert”; another could qualify the nature of the event with codes to indicate whether the event is “open to the public” or “private”. <pre><related qcode="eventrel:meeting" /></pre> ❖ a URI <pre><related uri="http://www.example.com/meeting" /></pre> ❖ a set of attributes to represent some arbitrary value. <pre><related rel="csrel:admission" value="7.0" valueunit="iso4217:EUR" value datatype="xs:decimal" /></pre>
<note>	Block	<p>A repeatable natural language note of additional information about the event, with an optional @role:</p> <pre><note role="nrole:toeditors"> Note to editors: an embargoed press release of the minutes of the meeting will be released by the COI within two weeks. </note> <note role="nrole:general"> The meeting was delayed by two days due to illness. </note></pre>

15.4.2 Event Relationships

Large news events may be split into a series of smaller, manageable events, arranged in a hierarchy which can express parent-child and peer relationships.

A “master” event may be notionally split into sub-events, which in turn may be split into further sub-events without limit. Each event instance can be managed separately yet handled and conveyed within the context of the larger realm of events of which it forms a part using the concept relationships <broadener>, <narrower>, <related> and <sameAs> relationship properties.

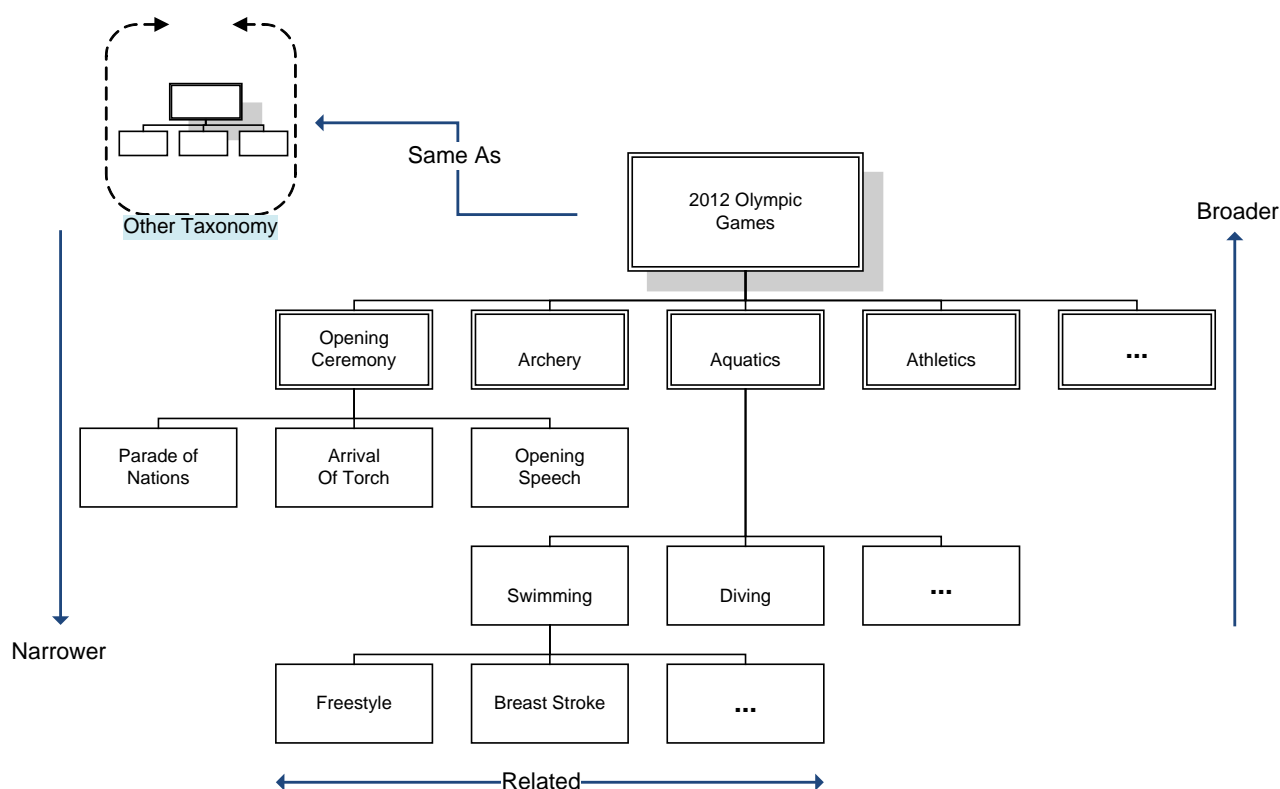


Figure 15: Hierarchy of Events created using Event Relationships

It is important to note from the above diagram that Broader, Narrower and Same As have specific relationships to the same type of Concept that is in this case Events. Related has no such restriction. thus:

- ❖ Aquatics is Broader than Breast Stroke or Swimming or Diving, but not Opening Speech.
- ❖ Diving is Narrower than Aquatics, but not Archery.
- ❖ Aquatics may be the Same As Aquatics in some other taxonomy, but not necessarily the Same As Swimming or Diving in some other taxonomy.
- ❖ Breast Stroke may be Related to an Event, a Person or other Concept Type.

The examples below show the use of the event relationship properties for the fictional Economic Policy Committee event:

Property	Type	Notes/Example
<broader>	Flexible Property (CCL)/Related Concept (PCL)	Repeatable. The event may be part of another event, in which case this can be denoted by <broader>. <pre> <broader type="cpnat: event" qcode="events: TR2012- 34625"> <name>Olympic Swimming Gala</name> </broader> </pre>
<narrower>	Flexible Property/ Related Concept (PCL)	Repeatable. May be used to indicate that the event has related child events. In this case we want to notify the receiver that a child of this event is the scheduled medal ceremony. <pre> <narrower qcode="events: TR2012- 34593"> <name>Men's 100m Freestyle Medal Ceremony</name> </narrower> </pre>
<related>	Flexible Property/ Related Concept (PCL)	Repeatable. May be used to denote a relationship to another concept or event. In this case, we want to link the event to an organisation which is not a participant, but may later form part of the coverage of the event. We qualify the relationship using @rel and the IPTC Item Relation <pre> <related @rel="IPTC: 34625" qcode="events: TR2012- 34625"> <name>Economic Policy Committee</name> </related> </pre>

Property	Type	Notes/Example
		<p>NewsCode:</p> <pre><related rel="irel:seeAlso" type="cpnat:org" qcode="org:asa"> <name>Amateur Swimming association</name> </related></pre>
<sameAs>	Flexible Property/ Related Concept (PCL)	<p>Repeatable. May be used to denote that this event is the same event in another taxonomy, for example another governing body's taxonomy of events:</p> <pre><sameAs qcode="iocevent:xxxxxx"> <name>Men's 100m Freestyle</name> </sameAs></pre>

15.4.3 Event Details Group

The first set of properties of Event Details is date-time information as described in [15.4.3.1](#) below. The further properties are described in [15.4.3.2](#) below

15.4.3.1 Dates and times

The IPTC intends that event dates and times in G2 align with the [iCalendar](#) (iCal) specification. This does not mean that dates and times are expressed in the same FORMAT as iCalendar, but the implementation of G2 properties that match iCalendar properties should be as set forth in the iCalendar specification.

The end dates/times of events are *non-inclusive*. Therefore a one day event on September 14, 2011 would have a *start date* (2011-09-14) and would have EITHER an *end date* of September 15 (2011-09-15) OR a *duration* of one day (syntax: see in table below).

The G2 syntax for expressing start and end times of events is a valid calendar date with optional time and offset; the following are valid:

- ❖ 2011-09-22T22:32:00Z (UTC)
- ❖ 2011-09-22T22:32:00-0500
- ❖ 2011-09-22

The following are NOT valid:

- ❖ 2011-09 (incomplete)
- ❖ 2011-09-22T22:32Z (if time is used, all parts MUST be present.
- ❖ 2011-09-22T22:32:00 (if time is used, time zone MUST be present.

When specifying the duration of an event, the date-part values permitted by iCalendar are W(eeks) and D(ays). In XML Schema, the only permitted date-part values are Y(ears), M(onths) and D(ays). (The permitted time parts H(ours), M(inutes) and S(econds) are the same in both XML Schema and iCalendar.)

The following table shows the permitted values for date-part in both standards

Duration	XML Schema	iCalendar
D(ays)	✓	✓
W(eeks)	✗	✓
M(onths)	✓	✗
Y(ears)	✓	✗

Since G2 uses XML Schema Duration, ONLY the values listed as permitted under the XML Schema column can be used, and therefore W(eeks) MUST NOT be used.



In addition the IPTC recommends that to promote inter-operability with applications that use the iCalendar standard, Y(ears) or M(onths) SHOULD NOT be used; only D(ays) should be used for the date part of duration values.

Duration units can be combined, in descending order from left to right, e.g.:

- ❖ P2D (duration of 2 days)
- ❖ P1DT3H (1 day and 3 hours – note the “T” separator)

iCalendar specifies that if the start date of an event is expressed as a Calendar date with no time element, then the end date MUST be to the *same* scale (i.e. no time) or if using duration, the ONLY permitted values are W(eeks) or D(ays). In this case, only D(ays) may be used in G2.

The following table lists the date-time properties of events details:

Property	Type	Notes/Example
<start>	Approximate Date Time	<p>Mandatory, non-repeatable property has two optional attributes, @approxstart and @approxend.</p> <p>The date part may be truncated, starting on the right (days) according to the precision required, but MUST, at minimum, have a year, for example:</p> <pre><start>2011- 06- 12T12: 30: 00Z</start></pre> <p>or</p> <pre><start>2011- 06- 12</start></pre> <p>If used, the time must be present in full, with time zone, and ONLY in the presence of the full date).</p> <p>The value of <start> expresses the precise date-time of the start of the event. With the information available, this might be a “best guess”.</p> <p>By using @approxstart and @approxend it is possible to qualify the start date-time by indicating the range of date-times within which the start will fall. (Note: these are NOT the approximate start and end of the <i>event</i> itself, only the range of <i>start date-times</i>)</p> <p>@approxstart indicates the start of the range. If used on its own, the end of the range of dates is the date-time value of <start></p> <p>For example, a possible start of an event on June 12, 2011, not before June 11, 2011, and no later than June 14 2011, would be expressed as:</p> <pre><start approxstart="2011- 06- 11" approxend="2011- 06- 14"> 2011- 06- 12 </start></pre>
<end>	Approximate Date Time	<p>Non-repeatable element to indicate the <i>non-inclusive</i> end time of the event, and optionally a range of values in which it may fall, using the same property type and syntax as for <start></p> <p>The <dates> wrapper may contain either an <end> date, or a <duration> but not both. A non-inclusive <end> date means, for example, that a one-day event starting on November 12, 2012 would have an end date of November 13, 2012.</p>
<duration>	XML	Non-repeatable. The time period during which the event takes place is



Property	Type	Notes/Example
	Schema Duration	<p>expressed in the form:</p> <p>PnYnMnDTnHnMnS</p> <p>P indicates the Period (required)</p> <p>nY = number of Years*</p> <p>nM = number of Months*</p> <p>nD = number of Days</p> <p>T indicates the start of the Time period (required if a time part is specified)</p> <p>nH = number of Hours</p> <p>nM = number of Minutes</p> <p>nS = number of Seconds</p> <p>* The IPTC recommends that these units should NOT be used.</p> <p>Example:</p> <pre><duration>PT3H</duration></pre> <p>The event will last for three hours. Note use of the "T" time separator even though no Date part is present.</p> <p>The <dates> wrapper may contain either an <end> date, or a <duration> but not both.</p>
<confirmation>	QCode	<p>Optional, non-repeatable. A QCode indicating whether the date-times for the event are confirmed or subject to possible change. The recommended IPTC NewsCode scheme currently has three possible values:</p> <pre><confirmation qcode="edconf: bothApprox" /></pre> <p>Indicates that both start and end dates are currently approximate or undefined</p> <pre><confirmation qcode="edconf: bothOk" /></pre> <p>Both start and end dates are confirmed</p> <pre><confirmation qcode="edconf: startApprox" /></pre> <p>The start date is approximate but the end date is confirmed</p>

15.4.3.1.1 Recurrence Properties

This is a group of optional properties that may be used to specify the complete set of recurring instances of an event, and conforms to the iCalendar specification, including the use of the same enumerated values for properties such as Frequency (@freq). Recurrence MUST be expressed using EITHER

<rDate>: one or more explicit date-times that the event is repeated, OR

<rRule>: one or more rules of recurrence.

Property	Type	Notes/Example
<rDate>	Date with optional Time	<p>Recurrence Date. Repeatable. If the recurrence occurs on a specific date, with an optional time part, or on several specific dates and times.</p> <pre><rdate>2011-03-27T14:00:00Z</rdate> <rdate>2011-04-03T16:00:00Z</rdate></pre>
<rRule>	Recurrence Rule	<p>Repeatable. The property has a number of attributes that may be used to define the rules of recurrence for the event.</p> <p>The only mandatory attribute is @freq, an enumerated string denoting the frequency of recurrence.</p> <pre><rRule freq="MONTHLY" /></pre> <p>The enumerated values of @freq are:</p> <ul style="list-style-type: none"> ❖ YEARLY ❖ MONTHLY ❖ DAILY ❖ HOURLY ❖ PER MINUTE ❖ PER SECOND <p>@interval indicates how often the rule repeats as a positive integer. The default is "1" indicating that for example, an event with a frequency of DAILY is repeated EACH day. To repeat an event every four years, such as the Summer Olympics, the Frequency would be set to 'YEARLY' with an Interval of "4":</p> <pre><rRule freq="YEARLY" interval="4" /></pre> <p>@until sets a Date with optional Time <i>after</i> which the recurrence rule expires:</p> <pre><rRule freq="MONTHLY" until="2011-12-31" /></pre> <p>@count indicates the number of occurrences of the rule. For example, an event taking place daily for seven days would be expressed as:</p> <pre><rRule freq="DAILY" count="7" /></pre> <p>A group of @byxxx attributes (as per the iCalendar BYxxx properties) are evaluated after @freq and @interval to further determine the occurrences of an event: @bymonth, @byweekno, @byyearday, @bymonthday, @byday, @byhour, @byminute, @bysecond, @bysetpos.</p> <p>The following code and explanation is based on an example from the iCalendar Specification at http://www.ietf.org/rfc/rfc5545.txt</p>



Property	Type	Notes/Example
		<pre><start>2011-01-11T8:30:00Z</start> <rRule freq="YEARLY" interval="2" bymonth="1" byday="su" [upper/lowercase - match in text] byhour="8 9" byminute="30" /></pre> <p>First, the interval="2" would be applied to freq="YEARLY" to arrive at "every other year". Then, bymonth="1" would be applied to arrive at "every January, every other year". Then, byday="SU" would be applied to arrive at "every Sunday in January, every other year".</p> <p>Then, byhour="8 9" (note that all multiple values are space separated) would be applied to arrive at "every Sunday in January at 8am and 9am, every other year". Then, byminute="30" would be applied to arrive at "every Sunday in January at 8:30am and 9:30am, every other year". Then, lacking information from rRule, the second is derived from <start>, to end up in "every Sunday in January at 8:30:00am and 9:30:00am, every other year".</p> <p>Similarly, if any of the @byminute, @byhour, @byday, @bymonthday or @bymonth rule part were missing, the appropriate minute, hour, day or month would have been retrieved from the <start> property.</p> <p>The @bysetpos attribute contains a non-zero integer "n" between -366 and 366 to specify the nth occurrence within a set of events specified by the rule. Multiple values are space separated. It can only be used with other @by* attributes.</p> <p>For example, a rule specifying monthly on any working day would be</p> <pre><rRule freq="MONTHLY" byday="MO TU WE TH FR" /></pre> <p>The same rule to specify the last working day of the month would be</p> <pre><rRule freq="MONTHLY" byday="MO TU WE TH FR" bypos="- 1" /></pre> <p>@wkst indicates the day on which the working week starts using enumerated values corresponding to the first two letters of the days of the week in English, for example "MO" (Monday), SA (Saturday), as specified by iCalendar.</p> <pre><rRule freq="WEEKLY" wkst="MO" /></pre>
<exDate>	Date with optional	Excluded Date of Recurrence. An explicit Date or Dates, with optional Time, excluded from the Recurrence rule. For example, if a regular



Property	Type	Notes/Example
	Time	<p>monthly meeting coincides with public holidays, these can be excluded from the recurrence set using <exDate></p> <pre><rRule freq="MONTHLY" until="2011-12-31" /> <exDate> 2011-04-06 </exDate></pre>
<exRule>	Recurrence Rule	<p>Excluded Recurrence Rule. The same attributes as <rRule> may be used to create a rule for excluding dates from a recurring series of events. For example, a regular weekly meeting may be suspended during the summer.</p> <pre><rRule freq="WEEKLY" until="2011-07-23" /> <rRule freq="WEEKLY" until="2011-12-24" /> <exRule freq="WEEKLY" until="2011-09-03" /></pre> <p>Note the order of the above statement: the <rRule> elements must come before <exRule></p> <p>The meaning being expressed is:</p> <p>"The event occurs weekly until Dec 24, 2011 with a break from after July 23, 2011 until September 3, 2011.</p>

15.4.3.2 Further Properties of Event Details

The event details group are wrapped by the <eventDetails> element

Property	Type	Notes/Example
<occurStatus>	QCode	<p>Optional, non-repeatable property to indicate the provider's confidence that the event will occur. The IPTC Event Occurrence Status NewsCode scheme</p> <p>http://cv.iptc.org/newscodes/eventoccurstatus/</p> <p>has values from "eos0" = "unplanned" though "eos5" = "certain to occur" indicating the provider's confidence of the status of the event.</p> <pre><occurStatus qcode="eocstat:eos5" /></pre>
<newsCoverageStatus>	Qualified Property	<p>Optional, non-repeatable element to indicate the status of planned news coverage of the event by the provider, using a QCode and (optional) <name> child element:</p> <pre><newsCoverageStatus qcode="ncstat:int"> <name></pre>



Property	Type	Notes/Example
		<p>Coverage Intended </name> </newsCoverageStatus></p>
<registration>	Block	<p>Optional, repeatable indicator of any registration details required for the event:</p> <pre><registration> Register online at
 http://www.example.com/registration.aspx/
 </registration></pre> <p>The property optionally takes a @role attribute. IPTC Registration Role NewsCode http://cv.iptc.org/newscodes/eventregrole/ may be used, which currently has four values:</p> <ul style="list-style-type: none"> ❖ Exhibitor ❖ Media ❖ Public ❖ Student <pre><registration role="eregrol:exhibReg"> Exhibitors must register online at
 http://www.example.com/exhibitor/register.aspx/ before May1, 2011
 </registration> <registration role="eregrol:pubReg"> The public may pre-register online at
 http://www.example.com/public/register.aspx/ to receive a special bonus pack.
 </registration></pre>
<accessStatus>	QCode	<p>Optional, repeatable property indicating the accessibility, the ease (or otherwise) of gaining physical access to the event, for example, whether easy, restricted, difficult. The QCodes represent a CV that would define these terms in more detail. For example, "difficult: may be defined as "Access includes stairways with no lift or ramp available. It will not be possible to install bulky or heavy equipment that cannot be safely carried by one person".</p> <pre><access qcode="access: easy" /></pre>
<subject>	Flexible Property	<p>Optional, repeatable. The subject classification(s) of the event, for example, using the IPTC Subject NewsCodes:</p> <pre><subject type="cpnat: abstract" qcode="medtop: 04000000"> <name xml:lang="en-GB"> Economy, Business and Finance </name> </subject> <subject type="cpnat: abstract" qcode="medtop: 20000271"> <name xml:lang="en-GB"></pre>



Property	Type	Notes/Example
		<pre> Financial and Business Service </name> </subj ect> <subj ect type="cpnat: abstract" qcode="medtop: 20000274"> <name xml:lang="en-GB"> Banki ng </name> </subj ect> </pre>
<location>	Flexible Property/ Flexible Location Property (PCL)	<p>Repeatable property indicating the location of the event with an optional <name>.</p> <pre> <locati on type="cpnat: poi "> <name>The Bank of Engl and, Threadneedl e Street, London, EC2R 8AH, UK</name> </locati on> </pre> <p>At PCL, a rich Concept-style structure may be used. (See the G2 Specification document for at www.newsml-g2.org for details)</p>
<participant>	Flexible Property/ Flexible Party Property (PCL)	<p>Optional, repeatable, The people and/or organisations taking part in the event. The type of participant is identified by @type and a QCode. The following example indicates a person, an organisation would be indicated (using the IPTC NewsCode) by type="cpnat:org"</p> <pre> <parti ci pant type="cpnat: person" qcode="pers: 32965"> <name xml:lang="en">Paul Tucker</name> </parti ci pant> </pre> <p>An IPTC Event Participant Role NewsCode is available: http://cv.iptc.org/newscodes/eventparticipantrole/ that holds roles such as "moderator", "director", "presenter"</p>
<participationRequirement>	Flexible Property	<p>Optional, repeatable element for expressing any required conditions for participation in, or attendance at, the event, expressed by a URI or QCode.</p> <pre> <parti ci pati onRequi rement qcode="partreq: accredited"> <name>Accredi tati on requi red</name> </parti ci pant> </pre>
<organiser>	Flexible Property/ Flexible Party Property (PCL)	<p>Optional, repeatable. Describes the organiser of the event.</p> <pre> <organi ser type="cpnat: org" uri="http://www.i ptc. org/"> <name xml:lang="en"> Internati onal Press Telecomuni cati ons Counci l </name> <name xml -l ang="fr"> </pre>



Property	Type	Notes/Example
		<pre>Comité International de Télécommunications de Presse </name> </organiser></pre> <p>The IPTC Event Organiser Role NewsCode, viewable at http://cv.iptc.org/newscodes/eventorganiserrole/ lists types of organiser such as, “venue organiser”, “general organiser”, “technical organiser”.</p>
<contactInfo>	Wrapper element	Indicates how to get in contact with the event. This may be a web site, or a temporary office established for the event, not necessarily the organiser or any participant. See note 15.4.3.2.1 below
<language>	-	<p>Optional, repeatable element describes the language(s) associated with the event using @tag with values that must conform to the IETF's BCP 47. An optional child element <name> may be added.</p> <pre><language tag="en"> <name>English</name> </language></pre>
<newsCoverage>		<p>DEPRECATED, Should not be used with Items conforming to EventsML-G2 1.6 and later.</p> <p>News Coverage is now part of the G2 Planning Item. (See Editorial Planning – the Planning Item)</p>

15.4.3.2.1 Contact Information

Contact information associated solely with the event, not any organiser or participant. For example, events often have a special web site and an event office which is independent of the organisers' permanent web site or office address.

The <contactInfo> element wraps a structure with the properties outlined below. An event may have many instances of <contactInfo>, each with @role indicating the purpose. These are controlled values, so a provider may create their own CV of address types if required, or use the IPTC Event Contact Info Role NewsCode, which has values of “general contact”, “media contact”, “ticketing contact”.

Each of the child elements of <contactInfo> may be repeated as often as needed to express different @roles, for example different “public” and “press” email addresses etc.

Property	Type	Notes/Example
<email>	Electronic Address	<p>An “Electronic Address” type allows the expression of @role (QCode) to qualify the information, for example</p> <pre><email role="addressrole: press"> office@iptc.org </email></pre>
<im>	Electronic Address	<pre><im role="imsrvc: reuters"> j.doe@iptc.org@reuters.net </im></pre>
<phone>	Electronic Address	Repeatable. Phone numbers should have a @role if necessary distinguish between different numbers:



Property	Type	Notes/Example
		<pre><phone role="phrole: public"> 1-123-456-7899 </phone> <phone role="phrole: press"> 1-123-456-7898 </phone></pre>
<fax>	Electronic Address	Fax number(s), as described for <phone> above.
<web>	IRI	<pre><web role="webrole: event"> www.iptc.org/springmeeting.html </web></pre>
<address>	Address	See 15.4.3.2.1.1 for details of Address properties. The Address may have a @role to denote the type of address it contains (e.g. office, personal) and may be repeated as required to express each address @role.
<note>	Block	Any other contact-related information, such as "Office closed for lunch from 12.30pm for three hours"

15.4.3.2.1.1 Address details <address>

The Address Type property may have a @role to indicate its purpose; the following table shows the available child properties. Apart from <line>, which is repeatable, each element may be used once for each <address>

Property	Type	Notes/Example
<line>	Internationalized string	As many as are needed
<locality>	Flexible Property	For example, name of a town or suburb. May be a URI, QCode, or no value with optional <name> child element
<area>	Flexible Property	For example, a county or region with optional <name> child element
<country>	Flexible Property	The country with optional <name> child element
<postalCode>	Internationalized String	The postal code, zip code or equivalent

For example:

```
<address role="addrole: registered">
  <line>20 Garrick Street</line>
  <locality>
    <name xml:lang="en">London</name>
  </locality>
  <country qcode="ISOCountryCode: UK">
    <name xml:lang="en">United Kingdom</name>
  </country>
  <postalCode>WC2E 9BT</postalCode>
</address>
```

15.5 Multiple Event Concepts in a Knowledge Item

Information about multiple events can be assembled in a Knowledge Item, which conveys a set of one or more Concepts. In this example, we will use a Knowledge Item to convey two related events.

The top-level <knowledgeItem> element contains identification, version and catalog information, in common with other G2 items:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<knowledgeItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20110126:qqwpi ruuew4712"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18">
```



```
conformance="power"
xml:lang="en">
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
<catalogRef href="http://www.example.com/events/event-catalog.xml" />
```

The <itemMeta> block conveys Management Metadata about the document and the <contentMeta> block can convey Administrative Metadata about the concepts being conveyed, and in this case we can show Descriptive Metadata that is common to all of the concepts in the Knowledge item,

The Core Descriptive Metadata properties common to all G2 Items that may be used in a Knowledge Item are: <language>, <keyword>, <subject>, <slugline>, <headline> and <description>.

```
<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2012-09-26T12:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
  <urgency>5</urgency>
  <contentCreated>2012-10-16T12:15:00Z</contentCreated>
  <contentModified>2012-10-16T14:35:00Z</contentModified>
  <subject qcode="medtop:20000304">
    <name>media</name>
  </subject>
  <subject qcode="medtop:20000309">
    <name>news agency</name>
  </subject>
  <subject qcode="medtop:20000763">
    <name>IT/computer sciences</name>
  </subject>
</contentMeta>
```

The two events in the listing are related, with the relationship indicated by the second event using the <broadener> property to show that it is an event which is part of the three-day event listed first

First event:

```
<conceptSet>
  <concept>
    <!-- FIRST EVENT! -->
    <conceptId created="2012-10-16T12:15:00Z" qcode="event:1234567" />
    <name>IPTC Autumn Meeting 2012</name>
    <eventDetails>
      ....
```

Second event: The IPTC News Content Working Party session below (event:91011123) has a 'broader' relationship to the IPTC Autumn Meeting above (event:1234567).

```
<concept>
  <!-- SECOND EVENT! -->
  <conceptId created="2012-09-30T12:00:00+00:00" qcode="event:91011123" />
  <name>Annomarket text analytics EU project</name>
  <broadener type="cpnat:event" qcode="event:1234567">
    <name>IPTC Autumn Meeting</name>
  </broadener>
  <eventDetails>
    ....
```

15.5.1 Full listing of the Event Knowledge Item

Note that although Knowledge Items are a convenient way to convey a set of related events, there is no requirement that all of the events in a KI must be related, or even that other concepts conveyed by the Knowledge Item are events; they may be people, organisations or other types of concept.

LISTING 15 Two Related Events in a Knowledge Item

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<knowledgeItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20101019:qqwpi ruuew4712"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18">
```



```

conformance="power"
xml:lang="en"
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-standards_25.xml" />
<catalogRef href="http://www.example.com/events/event-catalog.xml" />
<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2010-10-26T12:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
  <urgency>5</urgency>
  <contentCreated>2012-10-16T12:15:00Z</contentCreated>
  <contentModified>2012-10-16T14:35:00Z</contentModified>
  <subject qcode="medtop:20000304">
    <name>media</name>
  </subject>
  <subject qcode="medtop:20000309">
    <name>news agency</name>
  </subject>
  <subject qcode="medtop:20000763">
    <name>IT/computer sciences</name>
  </subject>
</contentMeta>
<conceptSet>
  <concept>
    <!-- FIRST EVENT! -->
    <!-- x -->
    <conceptId created="2012-10-16T12:15:00Z" qcode="event:1234567" />
    <name>IPTC Autumn Meeting 2012</name>
    <eventDetails>
      <dates>
        <start>2012-10-22T09:00:00-04:00</start>
        <duration>P3D</duration>
      </dates>
      <registration>Registration with the IPTC office is required.
        A <a href="http://www.example.com/register">
          web form</a> may be used until 02 October 2012
      </registration>
      <participationRequirement>
        <name>Membership</name>
        <definition>Only members of the IPTC and their invited
          guests may attend.
        </definition>
      </participationRequirement>
      <accessStatus qcode="accst:easy" />
      <language tag="en" />
      <organiser qcode="nprov:IPTC" role="orgrole:mainorganiser">
        <name>International Press Telecommunications Council</name>
        <organisationDetails>
          <founded>1965</founded>
        </organisationDetails>
      </organiser>
      <contactInfo>
        <email>mdirector@iptc.org</email>
        <note>Michael Steidl, Managing Director</note>
        <web>http://www.iptc.org</web>
      </contactInfo>
      <location>
        <name>Ria Novosti, Zubovsky Boulevard 4, Moscow, Russia</name>
        <related rel="frel:venue" qcode="ventyp:office" />
        <POIDetails>
          <position latitude="55.737126" longitude="37.591542" />
          <contactInfo>
            <web>http://en.rian.ru/</web>
          </contactInfo>
        </POIDetails>
      </location>
      <participant qcode="iptcofficer:SGuerillot">
        <name>Stéphane Guérillot</name>
        <definition role="drol:sessionrole">IPTC Chairman</definition>
      </participant>
      <participant qcode="iptcofficer:MSteidl">
        <name>Michael Steidl</name>
        <definition role="drol:sessionrole">Managing Director</definition>
      </participant>
    </eventDetails>
  </concept>

```




```
<concept>
  <!-- SECOND EVENT! -->
  <!-- x -->
  <conceptId created="2012-09-30T12:00:00+00:00" qcode="event:91011123" />
  <name>Annomarket text analytics EU project</name>
  <broader type="cpnat:event" qcode="event:1234567">
    <name>IPTC Autumn Meeting</name>
  </broader>
  <eventDetails>
    <dates>
      <start>2012-10-22T10:00:00-04:00</start>
      <duration>PT1H</duration>
    </dates>
    <participationRequirement>
      <name>Membership</name>
      <definition>Only members of the IPTC and their invited
        guests may attend
      </definition>
    </participationRequirement>
    <accessStatus qcode="accst:easy" />
    <language tag="en" />
    <participant qcode="iptcmember:JaredMcGinnis">
      <name>Jared McGinnis</name>
      <definition role="drol:sessionrole">Presenter</definition>
    </participant>
    <participant qcode="iptcofficer:MSteidl">
      <name>Michael Steidl</name>
      <definition role="drol:sessionrole">Moderator</definition>
    </participant>
  </eventDetails>
</concept>
</conceptSet>
</knowledgeItem>
```

15.5.2 Indicating changes to part of a Knowledge Item

When multiple Events are conveyed as Concepts in a Knowledge Item, and a sub-set of the Event Concepts is updated, providers may use the <partMeta> helper element to inform customers WHAT has changed in the new version of the Knowledge Item.

The <partMeta> element uses the standard XML ID/IDREF ; the example shows the Event with @id="eventA" was modified on '2010-09-15', as indicated by the Part Meta with @contentRefs="eventA" :

```
<knowledgeItem
  ...>
  <itemMeta>
    ...
  </itemMeta>
  <contentMeta>
    ...
  </contentMeta>
  <partMeta contentRefs="eventA">
    <contentModified>2010-09-15</contentModified>
  </partMeta>
  <conceptSet>
    <concept id="eventA">
      <!-- FIRST EVENT! -->
      <!-- x -->
      <conceptId created="2010-08-30T12:00:00+00:00" qcode="event:1234567" />
      ...
    </concept id="eventA">
    <concept>
      <!-- SECOND EVENT! -->
      <!-- x -->
      <conceptId created="2010-08-30T12:00:00+00:00" qcode="event:91011123" />
      ...
    </concept>
  </conceptSet>
</knowledgeItem>
```

15.6 Conveying Events in a NewsML-G2 Package

A news provider may wish to create a service that consists of collections of events that are significant to a specific editorial theme, for example, this could be the day's Top Finance Events.

When such Events are available as Concept Items, they may be referenced in a Package using `<itemRef>` and `@residref`. (See the [Quick Start - Packages](#) chapter for details of the Package Item structure and and Item references)

Alternatively, because they are managed and persistent Events identified using URIs, or their short format QCodes, collections of event concepts may be referenced within a Package `<group>` using Concept Reference `<conceptRef>` as follows:

```
<groupSet root="G1">
  <group id="G1" role="group: main">
    ...
    <conceptRef type="cinat: event" qcode="iptcevents: 20081007135637. 12">
      <name xml:lang="fr">Barack Obama arrive à Washington</name>
    </conceptRef>
    ...
  </group>
</groupSet>
```

Note the optional `@type` indicating that the referenced concept is an Event, and the optional `<name>` child element is a natural-language name for the event extracted from the concept being referenced.

Note the following guidelines on packaging Event Concepts:

- ❖ "Hint" properties – these are properties extracted from the original concept and conveyed with the reference to the concept – are restricted to `<name>` only when using `<conceptRef>`. When using `<itemRef>`, ANY property may be extracted from the referenced Concept Item.
- ❖ The provider must take care that "hint" properties do not re-define the original concept. When a provider distributes concepts using Concept Items or Knowledge Items, changes to the original concept MUST be issued via an updated Concept Item and/or Knowledge Item and if a `@version` is used with the `<itemRef>`, `@residref` must be updated to correspond with the revised concept's version.
- ❖ If for the referenced concept a Concept Item is available then this concept's Item MAY be referenced by `<itemRef>`; this is the preferred option. This MAY include hint properties copied from the Concept Item.
- ❖ If for the referenced concept a Concept Item is available then this concept MAY be referenced using `<conceptRef>` as a valid alternative option, for example when a package includes a mix of concepts for which some Concept Items are available and some are not available, but the provider wants to express the references in a consistent way. However, in a case such as this the constraint on the use of hints has to be accepted.
- ❖ If for a referenced concept there is no Concept Item and only a URI as identifier is available, then `<conceptRef>` MUST be used; no hint properties, except a name, are available as there is no Concept Item available as a source for copying hints.

The standard method for exchanging concept information in the NewsML-G2 context is the Knowledge Item, which is a container for concepts and the detailed properties of each concept being conveyed. A provider would use Knowledge Items to send comprehensive information about many events, and receivers might incorporate this information into their own editorial diary, of Day Book.

By contrast, when a Package is employed to reference Event concepts, this represents an Editorial product, in which a discrete number of events considered to be significant or pertinent to a particular topic, are selected and published.

The Events Package has the following characteristics:

- ❖ The value of the package is in the journalistic judgement used to select the events, not in the compilation of the event information itself.
- ❖ When using `<conceptRef>`, the package is a lightweight container that references each event and a human-readable name; no further details of the events are given. In this case, the package is

transient in nature: it will not be referenced over time except as an archived item of editorial content.

- ❖ It may be ordered to indicate the relative significance of the events in the context of the package, but no other relationship between the events is expressed or implied.
- ❖ New versions of the Package may be published; events may be replaced or deleted from the package, but details of the events themselves cannot be changed.

The characteristics of a KI are:

- ❖ It represents Knowledge; the value is in the compilation of the details about each event in the KI: the “when and where” and other practical information about each event.
- ❖ The KI contains persistent information about each event concept that may be expected to be frequently used and if necessary updated over time.
- ❖ The concepts in the KI cannot be ordered, but may be related.
- ❖ New versions of the KI can be published and the concepts that it contains can be updated, but they must not be deleted, as this would break existing references to the concepts.

15.7 Events Workflow

In the flow diagram shown below, the News Provider creates and compiles event information and publishes it in a Knowledge Item. The Customer’s events management system subscribes to this information.

Later, the News Provider’s editorial team reviews the events database and creates an ordered list of the “Top Ten” news events of the day. The News Provider publishes the list as a Package Item. The Package Item is ingested by the Customer’s editorial system, and the Customer’s journalists use it as a guide in planning the news coverage of the day, looking up details of each event (hosted in the Knowledge Item) as needed.

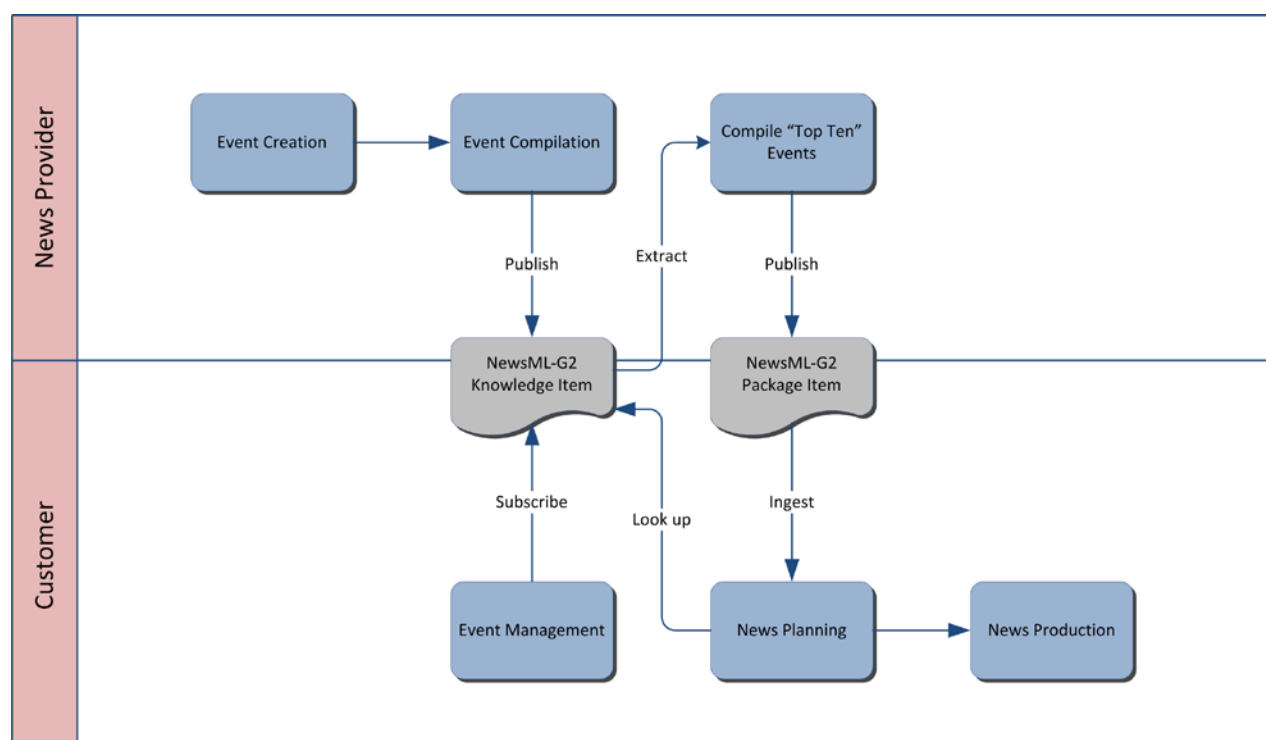


Figure 16: Event Flow using Package Items and Knowledge Items

15.8 Events in a News Item

Many news providers, particularly news agencies, provide their customers with event information as a list of events scheduled for coverage in a particular time frame, for example a list of cultural events taking place in a city’s theatres. In many cases, these were provided as a text story, with minimal mark-up.



Conveying these events in a structured way can be useful, as they are often shown as tables in print and on the Web, while there is little interest in making them persistent.

By using G2 properties, events such as these can be conveyed as a list capable of being machine-processed and formatted.

The Content Set of the Events News Item uses the <inlineXML> element to convey an <events> wrapper containing one or more <event> instances, each of which is a separate self-contained set of event information:

```
<contentSet>
  <inlineXML>
    <events>
      <event>
        <name>Annomarket text analytics EU project</name>
        <eventDetails>
          <dates>
            <start>2012-10-22T10:00:00-04:00</start>
            <end>2012-10-22T11:00:00-04:00</end>
          </dates>
        </eventDetails>
      </event>
    </events>
  </inlineXML>
</contentSet>
```

The full listing is shown below. Note the Item Class of this News Item is "ninat:composite".

LISTING 16 A Set of Events carried in a News Item

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20090122:qqwpiruuew4711"
  version="2" standard="NewsML-G2" standardversion="2.18" conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-standards_25.xml" />
  <itemMeta>
    <itemClass qcode="ninat:composite" />
    <provider qcode="nprov:IPTC" />
    <versionCreated>2012-10-18T12:44:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentSet>
    <inlineXML>
      <events>
        <!-- X -->
        <!-- FIRST EVENT! -->
        <!-- X -->
        <event>
          <name>IPTC Autumn Meeting</name>
          <eventDetails>
            <dates>
              <start>2012-10-22T09:00:00-04:00</start>
              <duration>P3D</duration>
            </dates>
            <location>
              <name>Ria Novosti, Zubovsky Boulevard 4, Moscow, Russia</name>
              <related rel="frel:venue" qcode="ventyp:office" />
              <POIDetails>
                <position latitude="55.737126" longitude="37.591542" />
                <contactInfo>
                  <web>http://en.rian.ru</web>
                </contactInfo>
              </POIDetails>
            </location>
          </eventDetails>
        </event>
        <!-- X -->
        <!-- SECOND EVENT! -->
        <!-- X -->
        <event>
          <name>Annomarket text analytics EU project</name>
          <eventDetails>
            <dates>
              <start>2012-10-22T10:00:00-04:00</start>
              <end>2012-10-22T11:00:00-04:00</end>
```

```

        </dates>
      </eventDetails>
    </event>
    <!--           X           -->
    <!--           THIRD EVENT!       -->
    <!--           X           -->
    <event>
      <name>Accidental Heroes</name>
      <definition>
        News stories and random incidents provide the inspiration behind
        this new production from the Lyric Young Company, which blends the
        inconsequential with the life-defining in a physical and visually
        arresting new show.
      <br />
        The Lyric Young Company has worked with award-winning
        writer/director Mark Murphy to create Accidental Heroes.
      <br />
      </definition>
      <related rel="frel:facility" qcode="facIncd:Food" />
      <related rel="frel:facility" qcode="facIncd:AirConditioning" />
      <eventDetails>
        <dates>
          <start>2012-10-23T19:30:00-04:00</start>
          <end>2011-02-05</end>
          <rRule freq="DAILY" byday="TH FR SA" />
        </dates>
      </eventDetails>
    </event>
  </events>
</inlineXML>
</contentSet>
</newsItem>

```

15.8.1 Adding event concept details to a News Item

Expressing Events as Concepts in a Concept Item gives implementers great scope for including and/or referencing events information in other Items, such as News Items conveying content. This is achieved by adding the event as a <subject> element of the <contentMeta> container. For example in **LISTING 15** an event was conveyed as a concept within the Concept Set of a Knowledge Item:

```

<knowledgeItem
...>
  <conceptSet>
    <concept>
      <!--           FIRST EVENT!       -->
      <!--           -->
      <conceptId created="2012-10-16T12:00:00+00:00" qcode="event:1234567" />
      <name>IPTC Autumn Meeting</name>
      <eventDetails ...
    </concept>
  </conceptSet>
</knowledgeItem>

```

The QCode **event:1234567** uniquely identifies the event, and later news coverage of the event can reference it using the <subject> element, as shown below:

```

<newsItem
...>
  <contentMeta>
    ...
    <subject type="cpnat:event" qcode="event:1234567">
      <name>IPTC Autumn Meeting 2012</name>
    </subject>
    ...
  </contentMeta>
  ...
</newsItem>

```

Referencing an event using Subject enables News Items to be searched and/or grouped by event, and also helps end-users manage the content for events with a wide coverage and many delivered Items.

15.8.2 Using <bag> to create a composite concept

Providers can put an event and related concepts together to create a new composite concept using a <bag> child element of <subject>.

For example, a financial news service sends a News Item conveying content about a takeover of a small company by a larger global company. This takeover story has an event concept, which enables the provider to add a subject property containing the QCode of the Event:

```
<subject type="cpnat: event" qcode="finevent: takeover123AB" />
```

... together with subject properties for each company:

```
<subject type="cpnat: organization" qcode="isin: Small Company" />
<subject type="cpnat: organization" qcode="isin: Global Company" />
```

However, a new, richer, composite concept combining the concept IDs for the Event *and* the two companies could be created instead using <bag>:

```
<newsItem ... >
...
<contentMeta>
...
<subject type="cpnat: abstract">
  <name>Global Company takes over Small Company</name>
  <bag>
    <bit type="cpnat: event" qcode="finevent: takeover123AB" />
    <bit type="cpnat: organization" qcode="isin: Small Company" />
    <bit type="cpnat: organization" qcode="isin: Global Company" />
  </bag>
</subject>
...
</contentMeta>
...
</newsItem>
```



Note that in this case, the <subject> does *not* have its own identifier; these properties must NOT be used when <subject> has a <bag> element

15.8.2.1 Using @significance with <bag>

An advantage of using the composite structure of <bag> is that implementers can “fine tune” the subject properties using @significance. This enables receiving applications to be more discerning on behalf of end-users, about how they filter and select news.

The significance of an event to all the concepts in a <bag> is not necessarily equal: for an end-user interested in news about a *small* company, the fact that it has been taken over by a large company is very significant. To the follower of news about the *large* company, the story may be less significant.

The code snippet below shows how the significance attribute would be used to refine the previous example, such that the event is more significant to the SmallCompany (100) compared to the GlobalCompany (10):

```
<subject type="cpnat: abstract">
  <name>Global Company takes over Small Company</name>
  <bag>
    <bit type="cpnat: event" qcode="finevent: takeover123AB" />
    <bit type="cpnat: organization" qcode="isin: Small Company" significance="100"/>
    <bit type="cpnat: organization" qcode="isin: Global Company" significance="10"/>
  </bag>
</subject>
```



Note: The <bag> MUST contain one <bit> of type “cpnat:event”
This feature was added from NewsML-G2 2.7 onwards. ONLY for the special case where the <bag> contains a <bit> referencing an event, may a @significance attribute be added to the other <bit> members of the <bag>.

16 Editorial Planning – the Planning Item

The news-gathering process is not an ad-hoc process. As outlined in [Chapter 6 How News Happens](#), professional news organisations need to be well-organised so that resources are used effectively, customers get the news they need, in time, and editorial standards and quality are maintained.

News planning information can improve collaboration, and thus efficiency, in B2B news exchange, and for some news agencies and their media customers, this has become an important area of development.

The Planning Item, added to the G2 family in NewsML-G2 v 2.7 and EventsML-G2 v1.6, addresses this need. As a “top level” NewsML-G2 Item, it is a sibling of News Item, Concept Item, Package Item and Knowledge Item.

The Planning Item now carries the News Coverage information that was previously expressed by Event Concept Items under the <eventDetails> wrapper, and has also been expanded to include additional features, such as the ability to track news deliverables against a previously-announced manifest of objects.

There are a number of advantages to the separation of Events and Planning information, including:

- ❖ News coverage information may change frequently in response to a news event. Previously, this would have meant re-sending all of the event information whenever the news planning changed. Now only the planning information is re-sent – a lighter processing overhead.
- ❖ News coverage is not always in response to a news event taking place at a specific place and time; it is also topic-based, such as “The best ski resorts for this winter”. Without the Planning Item, topic-based coverage such as this would require the creation of a “dummy” event as a placeholder for the coverage information. Now, no event information is required, merely a Planning Item that can carry the description of the topic and can be linked to the content once it is created.
- ❖ Organisations needed a way to announce planned coverage of an event, in terms of (say) a number of pictures, then inform receivers of progress and fulfilment. A Planning Item can provide a list of Items that have been delivered. The complementary “Delivery Of” property, which has been added as a property of other Items, can also link individual Items back to a Planning Item.
- ❖ Having news coverage information tightly-bound to an event caused issues when events span more than one news cycle. A single event can now have multiple Planning Items spanning different time periods.

Planning Items typically focus on the delivery of coverage for a single event or topic, but can be linked to other Planning Items to facilitate the coverage of more complex or long-term events.

The structure of a Planning Item is common to other NewsML-G2 Items:

- ❖ The top level <planningItem> properties of conformance, identification and versioning that we associate with “anyItem”
- ❖ Item Metadata contained in the <itemMeta> wrapper
- ❖ Administrative and Descriptive information related to the content wrapped in the <contentMeta> element.
- ❖ Special-purpose elements <assert> and <inlineRef> to group details of concepts and to provide references from content to metadata (PCL only) (see [25.2](#) and [25.3](#) for details on the use of these elements)
- ❖ Extension Points for providers to add properties from non-IPTC namespaces.
- ❖ A wrapper for the content payload, in this case a <newsCoverageSet>.

16.1 Item Metadata <itemMeta>

The standard properties of <itemMeta> are available for use in a Planning Item. The <itemClass> property uses a mandatory planning-specific IPTC “Planning Item Nature” NewsCode. The Scheme URI is: <http://cv.iptc.org/newscodes/plinature/> and recommended Scheme Alias is “plinat”:

```
| <itemClass qcode="plinat:newscoverage" />
```


16.2 Content Metadata <contentMeta>

The standard properties of the Administrative Metadata group may be used; a restricted set of Descriptive Metadata properties (the Core Descriptive Metadata Group, may be used. The Core Group is:
<language>, <keyword>, <subject>, <slugline>, <headline>, <description>

16.3 <newsCoverageSet>

The <newsCoverageSet> wraps one or more <newsCoverage> components (see [Figure 17](#)). Typically, each <newsCoverage> component is bound to each different class of Item to be delivered, i.e. a Planning Item for two texts, ten pictures and one graphic could have three <newsCoverage> components.

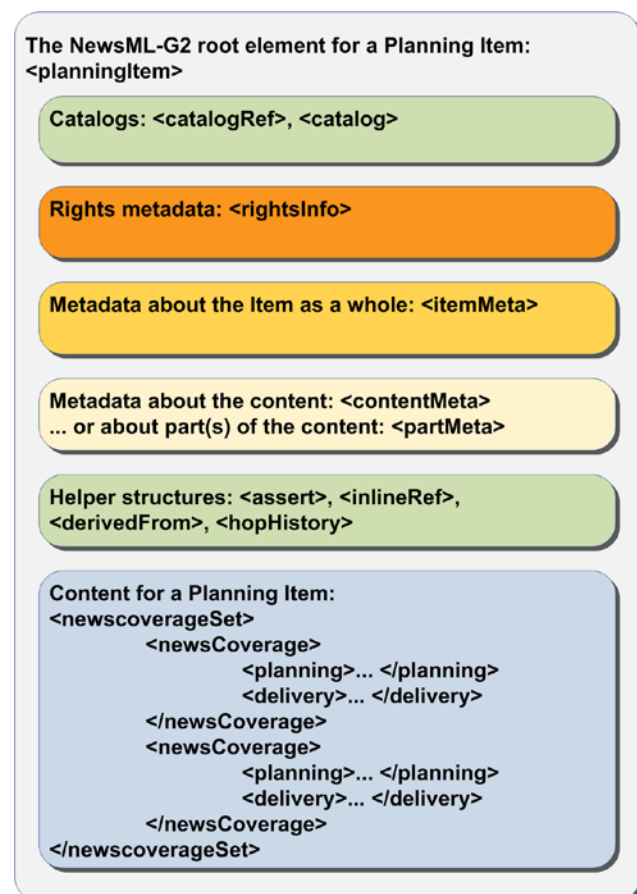


Figure 17: Planning Item: the newsCoverageSet has one or more newsCoverage components

16.4 <newsCoverage>

Each <newsCoverage> component MUST contain at least one <planning> element and optionally a <delivery> element.

16.5 <planning>

At CCL, planning information is expressed using a repeatable Block type <edNote> to give a natural language description of the planned coverage, optionally with some mark-up

`<edNote>Picture scheduled 2012- 11- 06T13: 00: 00+02: 00</edNote>`

Below is a complete example:

LISTING 17 Planning Item at CCL

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<planningItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="core">
```



```

xml:lang="en">
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
<catalogRef href="http://www.example.com/events/event-catalog.xml" />
<itemMeta>
  <itemClass qcode="plinat:newscoverage" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2012-10-18T13:05:00Z
</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
  <urgency>5</urgency>
  <contentCreated>2012-10-16T12:15:00Z</contentCreated>
  <contentModified>2012-10-16T13:35:00Z</contentModified>
</contentMeta>
  <newsCoverageSet>
    <newsCoverage>
      <planning>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage>
      <planning>
        <edNote>Picture scheduled 2012-11-06T13:00:00+02:00</edNote>
      </planning>
      <delivered>
        <deliveredItemRef>
          <guidref="tag:example.com, 2008:ART-ENT-SRV:USA20081220098658">
        </deliveredItemRef>
      </delivered>
    </newsCoverage>
  </newsCoverageSet>
</planningItem>

```

16.5.1 <planning> at Power Conformance Level

Using PCL capabilities for Planning can populate a customer's resource management applications with machine-readable information and detailed descriptive metadata, ready to be inherited by the arriving content, thus speeding up news handling and potentially increasing consistency and quality.

Used in this way, the Planning Item can bridge the workflows of provider and customer: the provider is seen as an available resource on the customer system with coverage information and news metadata capable of being updated in near real-time.

In NewsML-G2 version 2.18, two new properties were added that extend the information that can be conveyed by the <planning> child element of <newsCoverage>:

<newsContentCharacteristics> can be used to express comprehensive information about physical characteristics of the planned content, using the News Content Characteristics group of attributes. (see 13.9.9 *News Content Characteristics* in the NewsML-G2 Specification Document)⁹. The attributes include physical properties of still images, video and audio. In v2.18 the group was extended to add attributes for the line count of text content and the page count for content rendered as pages. An example is given in the table below.

<planningExtProperty> is an Extension Point for planning that permits the use of provider-specific properties while remaining within the G2 processing context. This is achieved by using an RDF-style subject-relationship-object expression.

Additional properties of <planning> at PCL:

Property	Type	Notes/Example
<g2ContentType>	String	Optional, non-repeatable element to indicate the MIME type of the intended coverage. The example below indicates that the

⁹ The NewsML-G2 Specification document available at www.newsml-g2.org is part of the download package for each Public Release of the standard, and contains a full list of attributes in the newsContentCharacteristics group.



Property	Type	Notes/Example
		<p>content to be delivered is a NewsML-G2 News Item.</p> <pre><g2ContentType> appl i cati on/vnd. i ptc. g2. newsi tem+xml </g2ContentType></pre>
<itemClass>	QCode	<p>Optional, non-repeatable element indicates the type of content to be delivered, using the IPTC News Item Nature NewsCode. Since the example will show a text article, the Item Class is “text”</p> <pre><i temCl ass qcode=“ni nat: text” /></pre>
<itemCount>	Empty	<p>The number of items to be delivered, expressed as a range:</p> <pre><i temCo unt rangefrom=“1” rangeto=“1” /></pre> <p>Both attributes @rangefrom (non-negative integer) and @rangeto (positive integer) MUST be used. The values are inclusive: rangefrom=“2” rangeto=“5” means 2 to 5 items will be delivered.</p>
<scheduled>	Approx Date Time	<p>Optional, non-repeatable. Indicates the scheduled time of delivery, and may be truncated if the precise date and time is not known. For example, if the content is scheduled to arrive at some unspecified time on a day, the value would be, for example:</p> <pre><schedul ed>2009- 10- 16</schedul ed></pre>
<service>	Qualified Property	<p>Optional, repeatable. The editorial service to which the content has been assigned by the provider and on which the receiver should expect to receive the planned content.</p> <pre><servi ce qcode=“srv: i ntwi re”> <name>International Wi re Servi ce</name> </servi ce></pre>
<assignedTo>	Flex1 Party Property	<p>The Flex1 Party Property Type extends the Flex Party Property Type by allowing a @role attribute to be a space separated list of QCodes.</p> <p><assignedTo> is an optional, non-repeatable element that holds the details of a person or organisation who has been assigned to create the announced content. It may hold as child elements any property from <personDetails> or <organisationDetails>, and properties from the Concept Definitions Group, and Concepts Relationships Group (see Concepts and Concept Items for details).</p> <p>The example shows the details of a person assigned to create the content:</p> <pre><assi gnedTo rol e=“erol e: edi tor” type=“cpnat: person” qcode=“pers: 54321” > <name>Sti lton Cheesewri ght</name> <personDetail s></pre>



Property	Type	Notes/Example
		<pre><contactInfo> <phone>1- 418- 4567</phone> <email>stilton@iptc.org</email> </contactInfo> </personDetails> </assignedTo></pre> <p>This information may be required internally by a news organisation as part of its event planning process, but perhaps may not be distributed to customers.</p> <p>Customers may be informed of the intended author/creator of planned coverage using the <by> property (see below)</p>
<newsContentCharacteristics>	Empty	<p>An element that enables providers to express physical properties of the planned item using attributes from the News Content Characteristics group:</p> <ul style="list-style-type: none"> - linecount. The count of the number of lines in text content <pre><newsContentCharacteristics linecount="205" /></pre> - pagecount. The number of pages of the planned content. <pre><newsContentCharacteristics pagecount="4" /></pre>
<planningExtProperty>		<p>For example, the planned item has a proprietary content rating. The rating is expressed using @rel with a QCode indicating the nature of the proprietary property, a @value and a @valuedatatype:</p> <pre><planningExtProperty rel="mediarel:hasParentalAdvisory" value="6" valuedatatype="xs:positiveInteger"/></pre>

16.5.1.1 Descriptive Metadata for <planning>

Metadata "hints" may also be added to assist receivers in preparing for the planned item, using the Descriptive Metadata Properties group:

Property	Type	Notes/Example
<by>	Label	<p>Optional, repeatable. Natural language author/creator information:</p> <pre><by>By Stilton Cheesewright</by></pre>
<creditline>	String	<p>Optional, repeatable. A freeform expression of the credit(s) for the content:</p> <pre><creditline>Additional reporting by Bertram Wooster</creditline></pre>
<dateline>	Label	<p>Optional, repeatable. Natural language information traditionally placed at the start of a text by some news agencies, indicating the place and time that the content was created:</p> <pre><dateline> Totley Towers, January 27, 2014 (Reuters) </dateline></pre>



Property	Type	Notes/Example
<description>	Block	Optional, repeatable. A free form textual description of the intended news coverage, with minimal mark-up permitted. The optional @role may use the IPTC Description Role NewsCode, which currently has values of “caption” and “summary”: <pre><description role="drol:summary"> Stilton Cheesewright will report on the proceedings from the NewsCodes Working Party </description></pre>
<genre>	Flex1 Concept Property	Optional, repeatable. The nature of the journalistic content that is intended for the news coverage. May be expressed by a QCode or URI value, with optional @type. Child elements may be any from the Concept Definition Group and the Concept Relationships Group (see Concepts and Concept Items): <pre><genre qcode="mygenre:main"> <name>Main article </name> </genre></pre>
<headline>	Label	Optional, repeatable. Headline that will apply to the content: <pre><headline>NewsCodes Working Party</headline></pre>
<keyword>	String	Optional, repeatable. A freeform term to assist in indexing the content: <pre><keyword>IPTC</keyword></pre>
<language>	-	Optional, repeatable. The language of the intended coverage, May have a @role to inform the receiver of the use of the language. The IPTC Language Role NewsCode currently has two values, “Subtitle” and “Voice Over” that apply to video content. The language @tag MUST be expressed using IETF BCP 47 and may have a child element of <name>: <pre><language tag="en-GB"> <name>UK English</name> </language></pre>
<slugline>	Internationalised String	Optional, repeatable. May have a @role and a @separator which indicates the character used as a delimiter between words or tokens used in the slugline: <pre><slugline separator="-">US-AUTO-BAILOUT</slugline></pre>
<subject>	Flex1 Concept Property	Optional, repeatable. Indicates the subject matter of the intended coverage. Child elements may be any from the Concept Definition Group and the Concept Relationships Group (see Concepts and Concept Items):



Property	Type	Notes/Example
		<pre> <subject qcode="medtop: 20000304"> <name>medi a</name> </subject> <subject qcode="medtop: 20000309"> <name>news agency</name> </subject> <subject qcode="medtop: 20000763"> <name>IT/computer sciences</name> </subject> Or <subject qcode="medtop: 20000309"> <name>news agency</name> <broadener qcode="medtop: 20000304"> <name>Medi a</name> </broadener> </subject> <subject qcode="medtop: 20000763"> <name>IT/computer sciences</name> </subject> </pre>

LISTING 18 Planning Item at PCL

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<planningItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="plinat:newscoverage" />
    <provider qcode="nprov:IPTC" />
    <versionCreated>2012-10-18T13:05:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2012-10-16T12:15:00Z</contentCreated>
    <contentModified>2012-10-16T13:35:00Z</contentModified>
  </contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2012-09-26T13:19:11Z">
      <planning>
        <g2contentType>application/ni tf+xml</g2contentType>
        <itemClass qcode="ni nat: text">
          <scheduled>2012-12-25T12:30:00-05:00</scheduled>
          <headline>Robinson: Must preserve bailout funds</headline>
          <edNote>Text 250 words</edNote>
        </planning>
      </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2012-09-26T15:19:11Z">
      <planning>
        <g2contentType>image/jpeg</g2contentType>
        <itemClass qcode="ni nat: picture">
          <scheduled>2012-12-25T12:00:00-05:00</scheduled>
          <edNote>Picture will be Robinson at today's news conference</edNote>
        </planning>
      </newsCoverage>
    </newsCoverageSet>
</planningItem>

```

16.6 The <delivery> component

The optional <delivery> component tells the receiver which parts of the planned coverage has been delivered. The delivered item(s) are referenced by one or more <deliveredItemRef> elements, each of which points to a delivered Item.

The complementary <deliverableOf> property may be added to the <itemMeta> of the corresponding delivered Item. This enables receivers to track back from delivered content to a specific News Coverage component. Although providers should keep the item references synchronised, it provides a bi-directional method for receivers to track the deliverables of a Planning Item, for example, if a News Item is delivered before the associated Planning Item is updated.

16.7 Delivered Items - <deliveredItemRef>

A set of properties to identify, locate and describe the delivered Item.

16.7.1 Delivered Item Reference: Core Conformance

The permitted attributes of <deliveredItemRef> are listed in the table below. In addition, a child <title> element may be added as metadata “hint” to the receiver.

Property	Type	Notes/Example
@rel	QCode	Indicates the relationship between the current Planning Item and the target Item
@href	URL	Locator for the target resource: <code>href="http://example.com/2008-12-20/pictures/foo.jpg"</code>
@residref	String	The provider's identifier for the target resource, i.e. the @guid of the Item: <code>residref="tag:example.com,2008:PIX:F0020081220098658"</code>
@version	String	The version of the target resource: the @version of the Item
@contenttype	String	The MIME type of the target resource: <code>contenttype="image/jpeg"</code>
@format	QCode	A refinement of the MIME type, taken from a CV
@size	String	Indicates the size of the target resource in bytes: <code>size="3764418"</code>

LISTING 19 Planning Item with delivery at CCL

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<planningItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="core"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="plinat:newscoverage" />
    <provider qcode="nprov:IPTC" />
    <versionCreated>2012-10-18T13:05:00Z
    </versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
```




```
<contentMeta>
<urgency>5</urgency>
<contentCreated>2012-10-16T12:15:00Z</contentCreated>
<contentModified>2012-10-16T13:35:00Z</contentModified>
</contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2012-09-26T13:19:11Z">
      <planning>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2012-09-26T15:19:11Z">
      <planning>
        <edNote>Picture scheduled 2012-12-25T12:00:00-05:00</edNote>
      </planning>
      <delivery>
        <deliveredItemRef>
          <title>Henry Robinson pictured today in New York</title>
        </deliveredItemRef>
      </delivery>
    </newsCoverage>
  </newsCoverageSet>
</planningItem>
```

16.7.2 Delivered Item Reference: Power Conformance

Additional attributes of <deliveredItemRef> may be used at PCL:

Property	Type	Notes/Example
@persistentidref	String	Reference to an element inside the target resource bearing an @id attribute, whose value must be persistent for all versions, i.e. for its entire lifecycle.
@validfrom, @validto	DateOpt Time	Date range (with optional time) for which the Item Reference is valid: validto="2010-11-20T17:30:00Z"
@id	XML ID	A local identifier for the <deliveredItemRef>
@creator	QCode	The person, organisation responsible for creating or editing this <deliveredItemRef> (i.e. not the referenced Item)
@modified	DateOpt Time	The date with optional time that this <deliveredItemRef> was last changed
@xml:lang	BCP47	The language used for this <deliveredItemRef>
@dir	Enumeration	The directionality of text, either "ltr" or "rtl" (left-to-right or right-to-left)
@rank	Non-negative integer	The rank of the <deliveredItemRef> among others in the Planning Item

16.7.2.1 Hint and Extension Point

At PCL, child elements from the NAR namespace or any other namespace may optionally be added. When using elements from the NAR, follow the rules set out in [Hint and Extension Point](#)

LISTING 20 Planning Item with delivery at PCL

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<planningItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18">
```



```
conformance="power"
xml:lang="en">
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
<catalogRef href="http://www.example.com/events/event-catalog.xml" />
<itemMeta>
  <itemClass qcode="plinat:newscoverage" />
  <provider qcode="nprov:IPTC" />
  <versionCreated>2012-10-18T13:05:00Z
</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
<urgency>5</urgency>
<contentCreated>2012-10-16T12:15:00Z</contentCreated>
<contentModified>2012-10-16T13:35:00Z</contentModified>
</contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2012-09-26T13:19:11Z">
      <planning>
        <g2contentType>application/ni tf+xml</g2contentType>
        <itemClass qcode="ni nat: text" />
        <scheduled>2012-12-25T12:30:00-05:00</scheduled>
        <headline>Robinson: Must preserve bailout funds</headline>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2012-09-26T15:19:11Z">
      <planning>
        <g2contentType>image/j peg</g2contentType>
        <itemClass qcode="ni nat: picture" />
        <scheduled>2012-12-25T12:00:00-05:00</scheduled>
        <edNote>Picture will be Robinson at today's news conference</edNote>
      </planning>
      <delivery>
        <deliveredItemRef
          guidref="tag:example.com,2008:ART-ENT-SRV:USA20081220098658">
          <title>Henry Robinson pictured today in New York</title>
        </deliveredItemRef>
      </delivery>
    </newsCoverage>
  </newsCoverageSet>
</planningItem>
```

17 SportsML-G2

17.1 Introduction¹⁰

Sports fixtures and results have long been an important part of the output of news agencies and media organisations and this activity continues to grow in line with the increasing world-wide interest in sport.

Historically, providers had to balance the need for detailed information with the constraints imposed by having to deliver timely transmissions of data over the (then) low-speed data circuits available. This led to sparse plain text or field-delimited data feeds that required precise formatting and processing in order to produce the required output, which was driven by the space constraints of print media.

Today, the picture is very different. The rise of the Web combined with the emergence of a global sports industry has created a demand for more detailed results and statistics. Although timeliness remains a key priority, modern communications and processing power have removed many of the old restrictions.

The legacy formats are therefore no longer adequate, but if the response to modern needs is a proliferation of specialised data formats, this will over time make the exchange and application of sports data more costly and difficult.

SportsML-G2 is designed to offer a flexible, extensible framework built on the re-usable components of the G2 Family that can handle all types of sports information using standard technology, including:

- ❖ Schedules (fixtures)
- ❖ Results
- ❖ Multi-media sports news
- ❖ Standings (league tables, player order-of-merit, rankings etc)
- ❖ Team statistics, including actions such as “fumbles”, “tackles”
- ❖ Player statistics, including career statistics and play statistics.
- ❖ Match statistics, including “plays”, or “actions”
- ❖ Betting or wagering information
- ❖ Venue information, including weather

Rather than try to fit all sports into a single generic model, special add-in modules enable widely-differing sports, such as golf, baseball and motor racing, to be accommodated within the standard framework.

17.2 Business Advantages

The G2 data model is well suited to the exchange of sports information. By its nature, sport has many entities: teams, players, officials, leagues etc that are routinely stored as structured data and can be used, updated, and re-used over time. These can be exchanged as Concepts and Knowledge Items.

Sports Events may therefore be modelled as actions and relationships involving these known entities, and by coding these in XML, the full value of this information can be harnessed using standard technology:

- ❖ Information can be easily imported into databases
- ❖ Using XSLT (XML-based formatting scripts) information can be rendered for the Web, print, mobile etc.
- ❖ The information can be used directly by dynamic applications using Java or similar tools.
- ❖ Providers and receivers are not restricted in their choice of taxonomies. This is crucial in sport where value-added knowledge bases may be maintained by different data owners.
- ❖ Extension points allow the standard to be adapted and customised to special needs within the standard framework

Finally, by building expertise around a single, extensible standard, information owners, providers, and consumers can benefit from reduced costs, greater reliability and faster time-to-market for compelling applications based on the rich variety of sports statistics and content that is increasingly available.

¹⁰ This Guide refers to SportsML-G2 v2.1. A later version 2.2 is available. Please see <http://dev.iptc.org/SportsML-22-ChangesAdditions> for a summary of changes and links to further information about SportsML-G2

17.3 Structure

SportsML-G2 content is conveyed within NewsML-G2 Items, (that is News Items, Package Items, Concept Items, Knowledge Items and Planning Items).

The structure of a SportsML-G2 document conveying a single piece of content, or multiple renditions of the same content, matches that of a NewsML-G2 News Item, as shown in [Figure 18](#).

The `<newsItem>`, `<itemMeta>` and `<contentMeta>` components are implemented as discussed in [Quick Start: NewsML-G2 Basics](#), which readers are encouraged to study first. Although a full description of all of the many features of SportsML-G2 is beyond the scope of this Guide, this Chapter documents the implementation of specific SportsML-G2 features and shows by example how its use may be extended.

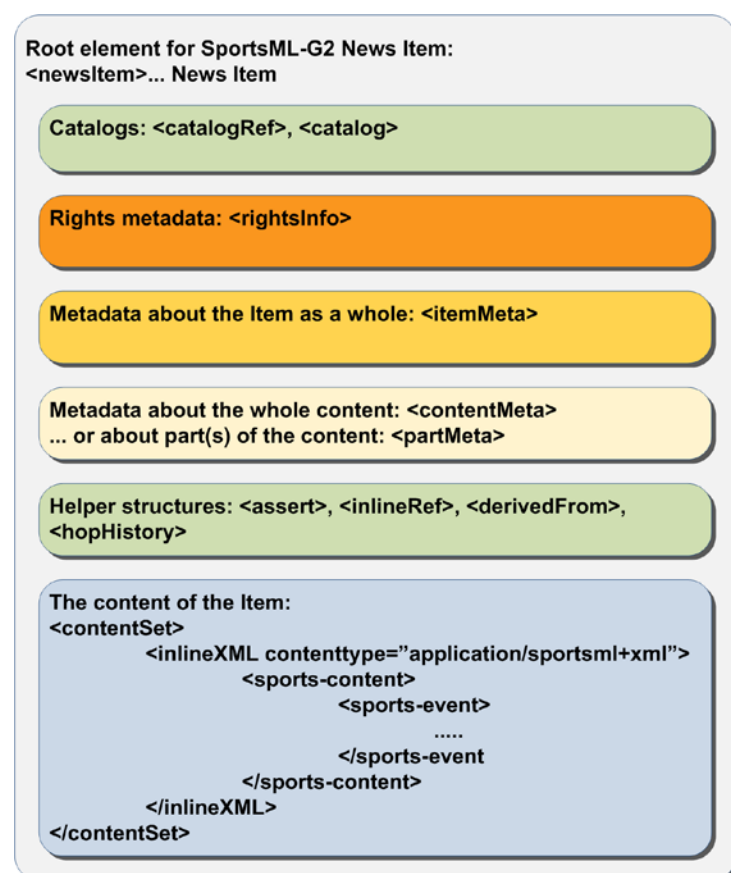


Figure 18: Structure of a SportsML-G2 document matches NewsML-G2

As shown in the diagram, the `<contentSet>` wrapper appropriate to SportsML-G2 is `<inlineXML>`, which wraps valid XML content.

17.3.1 XML Schema and Catalogs

The schema used by the `<newsItem>` wrapper is NewsML-G2, although the document is conveying SportsML-G2, because a NewsML-G2 `<newsItem>` is being used to convey the information. The SportsML-G2 namespace will be used in the scope of `<inlineXML>` wrapper of the document because it relates to the content.

The `@standard` is "NewsML-G2" and the appropriate `@version` value should reflect the version of NewsML-G2 being used, in the examples it is "2.18". If the Power Conformance Level is being used, the `@conformance` attribute MUST be present with a value of "power"

The Catalog references for a SportsML-G2 document are more likely to contain references to proprietary, or non-IPTC, catalogs. The IPTC Catalog, or those IPTC schemes that are mandatory, MUST be referenced. The IPTC Subject NewsCodes also have a comprehensive classification of sports, and the IPTC recommends their use as an aid to inter-operability.



Other provider-specific catalogs may be referenced. In the examples in this Chapter, we will use references to catalogs provided by kind permission of XML Team (www.xmlteam.com), an IPTC member and a specialist provider of sports information, who are a leading contributor to the SportsML community. Note that catalogs can be referenced using <catalogRef> and @href to a catalog file, and/or the <catalog> element may be used to reference the controlled vocabulary (scheme) using the <scheme> child element with @alias and @uri attributes.

```
<catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
<catalog>
  <scheme alias="fixture" uri="http://xmlteam.org/newscodes/fixture#" />
</catalog>
```

17.3.2 Item Metadata

A typical <itemMeta> block¹¹ may include the following:

```
<itemMeta>
  <itemClass qcode="ni nat: text" />
  <provider qcode="web: www.xmlteam.com" />
  <versionCreated>2009-02-02T11:17:00Z</versionCreated>
  <pubStatus qcode="stat: usable" />
  <fileName>xt.5932656-previ ew.xml</fileName>
</itemMeta>
```

Note that it contains a recommended <filename> for the item.

17.3.3 Content Metadata

In the following <contentMeta> block, the provider expresses the following facts about the content:

- ❖ When it was created
- ❖ Where the content was created
- ❖ Who created the content (note, not necessarily the provider of the SportsML-G2 document – see <provider> in Item Metadata)
- ❖ Alternative Identifiers for the content. Two will be given in the example, and the NewsML-G2 Specification recommends that they be distinguished using @type.
- ❖ The genre of the content, in this case a preview of a major league baseball fixture. In the example it is expressed more than once, using different controlled vocabularies.
- ❖ What, or whom, the content is about, including the fixture, teams and players, plus the classification of the sport (in this case using the IPTC Media Topic NewsCodes)

```
<contentMeta>
  <contentCreated>2000-02-02T12:17:00-05:00</contentCreated>
  <located qcode="ci ty: Philadelphia">
    <broadier qcode="reg: Pennsylvania" />
    <broadier qcode="cntry: USA" />
  </located>
  <creator qcode="web: sportsnetwork.com">
    <name>The Sports Network</name>
  </creator>
  <altId type="idtype: tsn-id" id="sportsnetwork.com-5932656" />
  <altId type="idtype: revision-id" id="1.mlb.com-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
  <genre type="spcpnat: spfixt" qcode="spfixt: pitcher-preview">
    <name xml:lang="en-US">Game Pitcher Preview</name>
    <broadier qcode="dclass: event-summary" />
  </genre>
  <genre type="spcpnat: dclass" qcode="dclass: event-summary" />
  <genre type="xts-genre: tsn-fixture" qcode="tsn-fixture: mlbprevi ewxml" />
  <language tag="en-US" />
  <subject qcode="medtop: 20000822">
    <name xml:lang="en-US">sport</name>
  </subject>
  <subject qcode="medtop: 20000849">
    <name xml:lang="en-US">Baseball</name>
    <broadier qcode="medtop: 20000822" />
  </subject>
```

¹¹ Implementers who are migrating from earlier versions of SportsML may download the SportsML 2.0:G2 Compliance Guide from the IPTC Web site www.iptc.org. This shows a comparison between pre-G2 structures and elements, and their G2 equivalents.



```
</subject>
<subject qcode="league:1.ml.b.com">
  <name xml:lang="en-US">Major League Baseball</name>
  <broader qcode="medtop:20000822" />
  <sameAs qcode="subj:15007000" />
</subject>
<subject type="spcpnat:conf" qcode="conf:1.ml.b.com-c.national">
  <name xml:lang="en-US">National</name>
  <broader qcode="league:1.ml.b.com" />
</subject>
<subject type="cpnat:event" qcode="event:1.ml.b.com-2007-e.19358" />
<subject type="spcpnat:team" qcode="team:1.ml.b.com-t.19">
  <name>Philadelphia Phillies</name>
</subject>
<subject type="spcpnat:team" qcode="team:1.ml.b.com-t.26">
  <name>Arizona Diamondbacks</name>
</subject>
<subject type="cpnat:person" qcode="person:1.ml.b.com-p.456">
  <name role="nrol:full">Doug Davis</name>
  <sameAs qcode="fssID:45679" />
</subject>
<subject type="cpnat:person" qcode="person:1.ml.b.com-p.123">
  <name role="nrol:full">Freddie Garcia</name>
  <sameAs qcode="fssID:45680" />
</subject>
<headline>Pitcher Preview: Arizona Diamondbacks (29-23) at Philadelphia
  Phillies (26-24), 7:05 p.m.</headline>
<slugline separator="-">AAV! PREVIEW-ARI-PHI</slugline>
</contentMeta>
```

17.3.4 InlineXML and the Sports Content component

Up to this point in the example, NewsML-G2 components have been re-used "as is" with minor variations. SportsML-G2 is a mark-up language specifically for sport, therefore as one might expect, it has structures and properties appropriate to those needs.

This becomes clear in the <contentSet> component. The content will be conveyed as SportsML-G2. The appropriate wrapper for this type of content is <inlineXML>.

```
<contentSet>
  <inlineXML contenttype="application/sportsml+xml">
    <sports-content xmlns="http://iptc.org/std/SportsML/2008-04-01/"
      xsi:schemaLocation="http://iptc.org/std/SportsML/2008-04-01/
        ../XSD/SportsML/sportsml-G2.xsd">
```

The Inline XML wrapper must contain a complete XML document from any namespace – in this case <sports-content> is the root element of a SportsML-G2 document and contains namespace and schema information.

The optional child elements <sports-metadata> and <article> are not documented here. The contents of <sports-metadata> have migrated to the <itemMeta> and <contentMeta> components, and <article> may be conveyed as a text using NewsML-G2.

It may be noted that the SportsML-G2 properties do not adhere to the NewsML-G2 convention of using "camel case" names (for example <sports-event> is used, not <sportsEvent>). This is to maintain backward compatibility with previous versions of SportsML, which existed as a standalone format before the creation of the G2 Standards.

Sports Content may hold the following child elements

17.3.4.1 Schedules or Fixtures <schedule>

A series of games, usually grouped by date. Also known as a fixture list.

17.3.4.2 Sports Event <sports-event>

Contains metadata and data about a sporting competition – a sporting content that generally ends with a winner

17.3.4.3 Standing <standing>

A series of team or individual records; facts about a team or player's performances during a match, competition, season or career,

17.3.4.4 Statistics, Tables <statistic>

A table comparing the performance of teams or players, such as a league table, order of merit, world ranking. Regular statistics may be linked to fixtures or competitions using a QCode.

17.3.4.5 Tournament <tournament>

A structured series of competitions within a sport that may be held over a period of time, for example, Rugby World Cup, UEFA Champions League, American Football Conference.

For the purposes of this example, we will discuss the properties of Sports Event

17.3.5 Sports Event wrapper in more detail

Sports Event MUST contain either a <player> OR a <team> child element. It SHOULD also contain an <event-metadata> OR <event-stats> child element.

Trying to fit different sports, for example golf and tennis, into a common structure would be impractical, if not impossible, without imposing severe limitations on the comprehensiveness of coverage. Using a plug-in model, SportsML-G2 can be extended to cover actions for virtually any sport.

Sports currently catered for are: American football, baseball, basketball, curling, golf, ice hockey, association football (aka soccer), motor racing, rugby football, tennis.

17.3.5.1 Event Metadata <event-metadata>

The contents of <event-metadata> are sports-specific: for example, for rugby, the time added to the match by the officials may be indicated.

Additionally, there are properties for recording the Event Sponsor and details of the Event Venue, The following information might apply to the Australian Open men's tennis final:

```
<event-metadata date-coverage-type="event"
  event-key="event:l.atp.com-2009-e.19358" event-status="mid-event"
  start-date-time="2009-01-28T19:05:00+11:00">
  <event-sponsor
    type="main"
    name="Kia Motors">
  </event-sponsor>
  <site>
    <site-metadata capacity="15,000" surface="Plexicushion"
      home-page-url="www.australianopen.com" site-key="site:aus209">
    <name>Rod Laver Arena, Melbourne Park</name>
    </site-metadata>
    <site-stats attendance="13,879" temperature="38"
      temperature-units="Centi grade" />
    </site>
  </event-metadata>
```

There is a choice of more than 40 attributes of <event-metadata> to record detailed information about an event, its timing, the venue, predicted weather – even its suitability for record-breaking on the day.

17.3.5.2 Event Statistics <event-stats>

Currently has a child element of <event-stats-motor-racing> which has attributes that enable the property to record summary events for a complete motor race such as laps covered, caution flags, margin of victory. For other sports, these statistics are carried at the <team> and <player> level.

17.3.5.3 Team <team>

Contains information about a team, its players, management, home location, and performance statistics. For example, a typical (not comprehensive <team> component:


```
<team>
  <team-metadata team-key="team:1.mlb.com-t.19" alignment="home">
    <name full="Philadelphia Phillies" role="nrol:full">Philadelphia Phillies</name>
  </team-metadata>
  <player>
    <player-metadata position-event="1" player-key="player:1.mlb.com-p.123">
      <name full="Freddie Garcia" role="nrol:full">
        Freddie Garcia
      </name>
      <name part="nprt:given">Freddie</name>
      <name part="nprt:family">Garcia</name>
    </player-metadata>
    <player-stats>
      <player-stats-baseball>
        <stats-baseball-pitching earned-runs="4.81" wins="1" losses="3"
          winning-percentage=".250"/>
      </player-stats-baseball>
    </player-stats>
  </player>
</team>
```

17.3.5.4 Player <player>

Contains information about sports competitors, their membership of teams, personal information and statistics. The example below shows how a player component might be used to record performance statistics in a rugby match:

```
<player>
  <player-metadata player-key="player:1.irb.org.worldcup-p.51" position-event="right-
  wing"
    uniform-number="14" status="starter">
    <name full="Vincent Clerc" />
  </player-metadata>
  <player-stats score="0">
    <player-stats-rugby>
      <stats-rugby-offensive tries-scored="0" conversion-attempts="0"
        conversions-scored="0"
        penalty-goal-attempts="0" penalty-goal-s-scored="0" drop-goal-s-scored="0"
        handling-errors="0" kicks-total="9" kicks-into-touch="5" runs="10"
        metres-gained="10"/>
      <stats-rugby-defensive tackles="10" tackles-missed="4"/>
      <stats-rugby-foul cautions-total="0" ejections-total="0"/>
    </player-stats-rugby>
  </player-stats>
</player>
```

17.3.5.5 Award <award>

A medal, cup, placing or other type of award, which can be assigned to an event, a team or a player using a local idref. .

```
<player id="P1">
  <player-metadata
    player-key="player:us.olympics.swim.123">
    <name full="Michael Phelps" />
  </player-metadata>
</player>
<award award-type="medal" name="Gold Medal"
  player-or-team-idref="P1" />
```

17.3.5.6 Event Actions <event-actions>

A container for recording play-by-play actions and scores specific to particular sports. Here is a possible action structure for a point in a tennis match:

```
<player id="P1">
  <player-metadata
    player-key="player:atp001">
    <name full="Roger Federer" />
  </player-metadata>
</player>
<player id="P2">
  <player-metadata
    player-key="player:atp:002">
    <name full="Rafal Nadal" />
  </player-metadata>
</player>
```

```

</player-metadata>
</player>
<event-actions id="T23564">
  <event-actions-tennis id="TS2G5P2">
    <action-tennis-point game="5">
      receiver-idref="P1" receiver-score="love"
      server-idref="P2" serve-number="first"
      server-score="30" set="2" win-type="unforced"
      winner-idref="P2">
    </action-tennis-point>
  </event-actions-tennis>
</event-actions>

```

17.3.5.7 Highlight <highlight>

A textual highlight for the event. An example of its use would be to provide a commentary-style text report on the progress of a match that could be used in a dynamic online application:

```

<highlight id="h. 718300" class="PRE_KICK-OFF">
  Liverpool make three changes from the side that drew with Wigan. Out go Babel, Lucas
  and Benayoun. In come Riera, Kuyt and Alonso. Robbie Keane meanwhile isn't even in the
  squad amid rumours that he is set to re-join Spurs.
  <sports-property formal-name="minutes-elapsed" value="0" />
</highlight>
<highlight id="h. 718351" class="KICK-OFF">
  We are underway here at a chilly Anfield.
  <sports-property formal-name="minutes-elapsed" value="0" />
</highlight>
<highlight id="h. 718358" class="COMMENT">
  Good counter attack from Liverpool as Fabio Aurelio overlaps down the left and plays
  it into Steven Gerrard, but Frank Lampard does well to track back and tackle his England
  team-mate.
  <sports-property formal-name="minutes-elapsed" value="2" />
</highlight>
....
<highlight id="h. 718402" class="YELLOW_CARD">
  Javier Mascherano is booked for a clumsy tackle on John Obi Mikel.
  <sports-property formal-name="minutes-elapsed" value="21" />
</highlight>
....
<highlight id="h. 718579" class="GOAL">
  LIVERPOOL 1-0 CHELSEA. Fabio Aurelio crosses from the left-hand side and TORRES dives
  in to head home and score a vital goal for Liverpool.
  <sports-property formal-name="minutes-elapsed" value="89" />
</highlight>

```

Note the use of @class to refine the semantic of each highlight property



17.3.5.8 Officials <officials>

A container for one or more officials taking part in the event. The child elements <official-metadata> and <official-stats> hold information about the official, including rating information. For example:

```
<officials>
  <official>
    <official-metadata official-key="official:fifa02345">
      position="referee">
        <name>Keith Burkinshaw</name>
      </official-metadata>
    <official-stats>
      <rating rating-issuer="FIFA" rating-maximum="8">
        rating-value="8" />
      </official-stats>
    </official>
  </officials>
```

17.3.5.9 Wagering Statistics <wagering-stats>

The betting on an event uses the <wagering-stats> component, which has possible child properties for expressing moneyline, odds, runline, straight spread and total score. The following example shows that the UK bookmaker William Hill is offering odds on an event of 10-1 from 8-1, together with a timestamp.

```
<wagering-stats comment="Multiple wagering-odds can be given">
  <wagering-odds bookmaker-key="book:uk23456">
    bookmaker-name="William Hill"
    date-time="20090202T15:00:00Z"
    numerator="10"
    denominator="1"
    numerator-opening="8"
    denominator-opening="1">
  </wagering-odds>
</wagering-stats>
```

LISTING 21 A complete sample SportsML-G2 document

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:xts="www.xmlteam.com"
  guid="tag:xmlteam.com,2009:xt.5932656-pitcher-preview"
  version="3"
  standard="NewsML-G2" standardversion="2.18"
  conformance="power" xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="inat:text" />
    <provider qcode="web:www.xmlteam.com" />
    <versionCreated>2010-10-19T16:00:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
    <fileName>xt.5932656-pitcher-preview.xml</fileName>
  </itemMeta>
  <contentMeta>
    <contentCreated>2000-02-02T12:17:00-05:00</contentCreated>
    <located qcode="city:Philadelphia">
      <broadier qcode="reg:Pennsylvania" />
      <broadier qcode="cntry:USA" />
    </located>
    <creator qcode="web:sportsnetwork.com">
      <name>The Sports Network</name>
    </creator>
    <altId type="idtype:tsn-id" id="sportsnetwork.com-5932656" />
    <altId type="idtype:revision-id"
      id="1.mlb.com-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
    <genre type="spcpnat:spfixt" qcode="spfixt:pitcher-preview">
      <name xml:lang="en-US">Game Pitcher Preview</name>
      <broadier qcode="dclass:event-summary" />
    </genre>
    <genre type="spcpnat:dclass" qcode="dclass:event-summary" />
    <genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />
    <language tag="en-US" />
    <subject qcode="medtop:20000822">
      <name xml:lang="en-US">sport</name>
    </subject>
    <subject qcode="medtop:20000849">
```



```

    <name xml:lang="en-US">Baseball</name>
    <broader qcode="medtop:20000822" />
</subject>
<subject qcode="league:1.mlb.com">
    <name xml:lang="en-US">Major League Baseball</name>
    <broader qcode="medtop:20000822" />
    <sameAs qcode="subj:15007000" />
</subject>
<subject type="spcpnat:conf" qcode="conf:1.mlb.com-c.national">
    <name xml:lang="en-US">National</name>
    <broader qcode="league:1.mlb.com" />
</subject>
<subject type="cpnat:event" qcode="event:1.mlb.com-2007-e.19358" />
<subject type="spcpnat:team" qcode="team:1.mlb.com-t.19">
    <name>Philadelphia Phillies</name>
</subject>
<subject type="spcpnat:team" qcode="team:1.mlb.com-t.26">
    <name>Arizona Diamondbacks</name>
</subject>
<subject type="cpnat:person" qcode="person:1.mlb.com-p.456">
    <name role="nrol:full">Doug Davis</name>
    <sameAs qcode="fssID:45679" />
</subject>
<subject type="cpnat:person" qcode="person:1.mlb.com-p.123">
    <name role="nrol:full">Freddy Garcia</name>
    <sameAs qcode="fssID:45680" />
</subject>
<headline>Pitcher Preview: Arizona Diamondbacks (29-23) at Philadelphia
    Phillies (26-24), 7:05 p.m.</headline>
<slugline separator="-">AAV! PREVIEW-ARI-PHI</slugline>
</contentMeta>
<contentSet>
    <inlineXML contentType="application/sportsml+xml">
        <sports-content xmlns="http://iptc.org/std/SportsML/2008-04-01/"
            xsi:schemaLocation="http://iptc.org/std/SportsML/2008-04-01/
                ../XSD/SportsML/sportsml-G2.xsd">
            <sports-event>
                <event-metadata date-coverage-type="event"
                    event-key="event:1.mlb.com-2009-e.19358"
                    date-coverage-value="1.mlb.com-2009-e.19358"
                    event-status="pre-event" start-date-time="2009-02-06T19:05:00-05:00">
                </event-metadata>
                <team>
                    <team-metadata team-key="team:1.mlb.com-t.26"
                        alignment="away">
                        <name full="Arizona Diamondbacks" role="nrol:full">
                            Arizona Diamondbacks</name>
                    </team-metadata>
                    <player>
                        <player-metadata position-event="1"
                            player-key="player:1.mlb.com-p.456">
                            <name full="Doug Davis" role="nrol:full">Doug Davis</name>
                            <name part="nprt:given">Doug</name>
                            <name part="nprt:family">Davis</name>
                        </player-metadata>
                        <player-stats>
                            <player-stats-baseball>
                                <stats-baseball-pitching earned-runs="3.57"
                                    wins="2" losses="6" winning-percentage=".250" />
                            </player-stats-baseball>
                        </player-stats>
                    </player>
                </team>
                <team>
                    <team-metadata team-key="team:1.mlb.com-t.19"
                        alignment="home">
                        <name full="Philadelphia Phillies" role="nrol:full">
                            Philadelphia Phillies</name>
                    </team-metadata>
                    <player>
                        <player-metadata position-event="1"
                            player-key="player:1.mlb.com-p.123">
                            <name full="Freddy Garcia" role="nrol:full">
                                Freddy Garcia</name>
                            <name part="nprt:given">Freddy</name>
                            <name part="nprt:family">Garcia</name>
                        </player-metadata>
                        <player-stats>
                            <player-stats-baseball>

```



```
<stats-baseball-pitching earned-runs="4.81"  
wins="1" losses="3" winning-percentage=".250" />  
</player-stats-baseball>  
</player-stats>  
</player>  
</team>  
</sports-event>  
</sports-content>  
</inlineXML>  
</contentSet>  
</newsItem>
```

17.3.6 SportsML-G2 Packages

To convey a package of SportsML-G2 items, simply use the NewsML-G2 <packageItem>, as covered in **Quick Start - Packages**. For example, we could send a text article to accompany the sports data document shown in the previous listing.

The text document is a News Industry Text Format (NITF¹²) document, conveyed in a NewsML-G2 News Item, as shown below.

LISTING 22 Sports story in NewsML-G2/NITF

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>  
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:xts="http://www.xmlteam.com/2009/xt.5932656-preview" version="3" standard="NewsML-G2" standardversion="2.18" conformance="power" xml:lang="en-US">  
  <catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />  
  <catalogRef href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />  
  <itemMeta>  
    <itemClass qcode="ninat:text" />  
    <provider qcode="web:www.xmlteam.com" />  
    <versionCreated>2010-10-19T11:45:00-04:00</versionCreated>  
    <pubStatus qcode="stat:usable" />  
    <fileName>xt.5932656-preview.xml</fileName>  
  </itemMeta>  
  <contentMeta>  
    <contentCreated>2009-02-02T11:15:00-05:00</contentCreated>  
    <located qcode="city:Philadelphia">  
      <broader qcode="reg:Pennsylvania" />  
      <broader qcode="cntry:USA" />  
    </located>  
    <creator qcode="web:sportsnetwork.com">  
      <name>The Sports Network</name>  
    </creator>  
    <altId type="idtype:tsn-id" id="sportsnetwork.com-5932656" />  
    <altId type="idtype:revision-id" id="1.mlb.com-2007-e.19358-pre-event-coverage-sportsnetwork.com" />  
    <genre type="spcpnat:spfxt" qcode="spfxt:pre-event-coverage">  
      <name xml:lang="en-US">Game Preview</name>  
      <broader qcode="dclass:event-summary" />  
    </genre>  
    <genre type="spcpnat:dclass" qcode="dclass:event-summary" />  
    <genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />  
    <language tag="en-US" />  
    <subject qcode="medtop:20000822">  
      <name xml:lang="en-US">sport</name>  
    </subject>  
    <subject qcode="medtop:20000849">  
      <name xml:lang="en-US">Baseball</name>  
      <broader qcode="medtop:20000822" />  
    </subject>  
    <subject qcode="league:1.mlb.com">  
      <name xml:lang="en-US">Major League Baseball</name>  
      <broader qcode="medtop:20000822" />  
      <sameAs qcode="subj:15007000" />  
    </subject>
```

¹² The IPTC's XML standard for marking up text.



```
<subject type="spcpnat: conf" qcode="conf: 1. ml b. com- c. national">
  <name xml: lang="en- US">National</name>
  <broader qcode="league: 1. ml b. com" />
</subject>
<subject type="cpnat: event" qcode="event: 1. ml b. com- 2007- e. 19358" />
<subject type="spcpnat: team" qcode="team: 1. ml b. com- t. 19">
  <name>Philadelphia Phillies</name>
</subject>
<subject type="spcpnat: team" qcode="team: 1. ml b. com- t. 26">
  <name>Arizona Diamondbacks</name>
</subject>
<subject type="cpnat: person" qcode="person: 1. ml b. com- p. 456">
  <name>Doug Davis</name>
  <sameAs qcode="fssID: 45679" />
</subject>
<subject type="cpnat: person" qcode="person: 1. ml b. com- p. 123">
  <name>Freddy Garcia</name>
  <sameAs qcode="fssID: 45680" />
</subject>
<headline>Preview: Arizona Diamondbacks (29-23) at Philadelphia Phillies
(26-24), 7: 05 p. m.</headline>
<slugline separator="-">AAV! PREVIEW- ARI - PHI</slugline>
<description role="drol: summary">A pair of teams coming off big weekend sweeps
will square off tonight at Citizens Bank Park, where the Philadelphia Phillies
welcome the Arizona Diamondbacks for the start of a three-game
series.</description>
</contentMeta>
<contentSet>
  <inlineXML contentType="application/nitf+xml">
    <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/"
      xsi:schemaLocation="http://iptc.org/std/NITF/2006-10-18/
        ../XSD/NITF/3.5/specification/nitf-3-5.xsd">
      <body>
        <body.head>
          <headline>
            <hl1>Arizona Diamondbacks (29-23) at Philadelphia Phillies (26-24),
              7: 05p. m.</hl1>
          </headline>
          <byline>
            <byttl>Sports Network</byttl>
          </byline>
          <abstract>
            <p>A pair of teams coming off big weekend sweeps will square off
              tonight at Citizens Bank Park, where the Philadelphia Phillies
              welcome the Arizona Diamondbacks for the start of a three-game
              series.</p>
          </abstract>
        </body.head>
        <body.content>
          <p>(Sports Network) - A pair of teams coming off big weekend sweeps
            will square off tonight at Citizens Bank Park, where the Philadelphia
            Phillies welcome the Arizona Diamondbacks for the start of a three-game
            series.</p>
          <p>The Phillies picked up some momentum with a three-game sweep of the
            Atlanta Braves, while the Diamondbacks captured all four of their games
            against the Houston Astros.</p>
          <p>On Sunday, Livan Hernandez threw his first complete game in almost
            two years to lead the Diamondbacks to an 8-4 win over the Astros to
            complete the sweep.</p>
          <p>Arizona sends Doug Davis to the hill tonight, and the lefty is 0-4
            over his last five starts. That includes losses in each of his last
            three outings.</p>
          <p>The Phillies, meanwhile, notched their first sweep of the year that
            culminated with Sunday's 13-6 pounding of the Braves. Ryan Howard
            flashed some of the power that won him the NL MVP last season, jacking
            a pair of two-run homers in the win.</p>
          <p>Right-hander Freddy Garcia takes the hill for the Phillies and is 1-
            3 with a 4.81 earned run average on the year. Since winning his first
            game in a Philadelphia uniform back on April 22, Garcia has gone 0-2
            over his last six starts.</p>
        </body.content>
      </body>
    </nitf>
  </inlineXML>
</contentSet>
</newsItem>
```

17.3.6.1 Referencing the packaged items

Each of the documents listed has a <filename> property in <itemMeta> that gives the recommended name that should be used when storing the document in the receiver's system. The data file's property is:

```
<fileName>xt. 5932656- pitcher- preview. xml </fileName>
```

The property for the text document is:

```
<fileName>xt. 5932656- preview. xml </fileName>
```

These filenames will be referenced by the NewsML-G2 Package Item.

17.3.6.2 Item Metadata and Content Metadata

The <itemMeta> contains Management Metadata about the Item as a whole, and some Content Metadata (located, creator, and subject) that is common to both items is repeated in the <contentMeta> of the package as an aid to processing by the receiver.

17.3.6.3 The Group Set <groupSet> component

These processing "hints" are also a feature of the item references in the package payload component <groupSet>. Some of the metadata from the referenced items is inherited by the <itemRef> property for each member of the package. This enables the receiver to carry out some processing of the package without needing to retrieve the referenced items, for example to display on a web summary page.

In order to achieve this, each item reference contains a @contenttype to alert the receiver to the format of the content being referenced, and further the genre, or nature, of the content, together with headline and summary.

This is a simple package of two related items, so no package hierarchy is needed. The package <groupSet> structure and its relationship to the information assets that it references may be illustrated in diagram form as shown in **Figure 19**

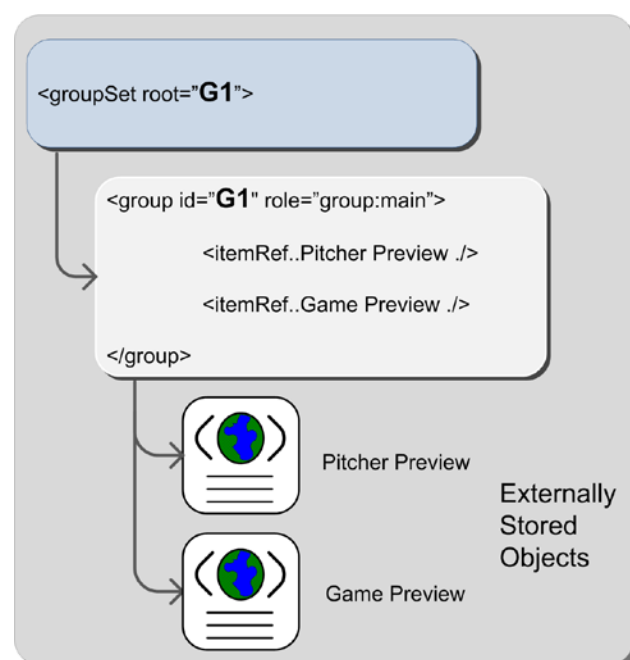


Figure 19: Diagram of Package Group structure

This produces a <groupSet> component in XML as follows:

```

<groupSet root="G1">
  <group id="G1" role="group:main">
    <itemRef href="xt. 5932656- pitcher- preview"
      contentType="application/sportsml+xml">
      <genre type="spcpnat:spfxt" qcode="spfxt: pitcher- preview">
        <name xml:lang="en-US">Game Pitcher Preview</name>
        <broader qcode="dclass: event- summary" />
      </genre>
    </itemRef>
    <itemRef href="xt. 5932656- preview"
      contentType="text/xml">
      <genre type="spcpnat:spfxt" qcode="spfxt: preview">
        <name xml:lang="en-US">Game Preview</name>
        <broader qcode="dclass: event- summary" />
      </genre>
    </itemRef>
  </group>
</groupSet>
  
```




```

<genre type="spcpnat:dclass" qcode="dclass:event-summary" />
<genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />
<headline>Pitcher Preview: Arizona Diamondbacks (29-23) at Philadelphia
  Phillies (26-24), 7:05 p.m.
</headline>
<description></description>
</itemRef>
<itemRef href="xt.5932656-preview" contentType="application/ni tf+xml">
  <genre type="spcpnat:spfixt" qcode="spfixt:pre-event-coverage">
    <name xml:lang="en-US">Game Preview</name>
    <broadener qcode="dclass:event-summary" />
  </genre>
  <genre type="spcpnat:dclass" qcode="dclass:event-summary" />
  <genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />
  <headline>Preview: Arizona Diamondbacks (29-23) at Philadelphia
    Phillies (26-24), 7:05 p.m.
  </headline>
  <description role="drol:summary">A pair of teams coming off big weekend
    sweeps will square off tonight at Citizens Bank Park, where the
    Philadelphia Phillies welcome the Arizona Diamondbacks for the start of a
    three-game series.
  </description>
</itemRef>
</group>
</groupSet>

```

The complete Package Listing is shown below.

LISTING 23 SportsML-G2 Package

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<packageItem xmlns="http://iptc.org/std/nar/2006-10-01/" xmlns:xts="www.xmlteam.com"
  guid="urn:iptc.org:20060101:sports1" version="3"
  conformance="power"
  standard="NewsML-G2"
  standardversion="2.18"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="ninat:composite" />
    <provider qcode="web:xmlteam.com" />
    <versionCreated>2010-10-19T15:00:09Z</versionCreated>
  </itemMeta>
  <contentMeta>
    <contentCreated>2009-02-02T11:45:00-05:00</contentCreated>
    <located qcode="city:Philadelphia">
      <broadener qcode="reg:Pennsylvania" />
      <broadener qcode="cntry:USA" />
    </located>
    <creator qcode="web:sportsnetwork.com">
      <name>The Sports Network</name>
    </creator>
    <altId type="idtype:tsn-id" id="sportsnetwork.com-5932656" />
    <altId type="idtype:revision-id"
      id="1.mlb.com-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
    <genre type="spcpnat:dclass" qcode="dclass:event-summary" />
    <genre qcode="xts-package-type:pre-event-coverage" />
    <language tag="en-US" />
    <subject qcode="subj:15000000">
      <name xml:lang="en-US">sport</name>
    </subject>
    <subject qcode="subj:15007000">
      <name xml:lang="en-US">Baseball</name>
      <broadener qcode="subj:15000000" />
    </subject>
    <subject qcode="league:1.mlb.com">
      <name xml:lang="en-US">Major League Baseball</name>
      <broadener qcode="subj:15007000" />
      <sameAs qcode="subj:15007001" />
    </subject>
    <subject type="spcpnat:conf" qcode="conf:1.mlb.com-c.national">
      <name xml:lang="en-US">National</name>
      <broadener qcode="league:1.mlb.com" />
    </subject>
    <subject type="cpnat:event" qcode="event:1.mlb.com-2007-e.19358" />
    <subject type="spcpnat:team" qcode="team:1.mlb.com-t.19">

```



```

    <name>Philadelphia Phillies</name>
  </subject>
  <subject type="spcpnat:team" qcode="team:1.ml.b.com-t.26">
    <name>Arizona Diamondbacks</name>
  </subject>
</contentMeta>
<groupSet root="G1">
  <group id="G1" role="group:main">
    <itemRef href="xt.5932656-pitcher-preview"
      contentType="application/sportsml+xml">
      <genre type="spcpnat:spfixt" qcode="spfixt:pitcher-preview">
        <name xml:lang="en-US">Game Pitcher Preview</name>
        <broader qcode="dclass:event-summary" />
      </genre>
      <genre type="spcpnat:dclass" qcode="dclass:event-summary" />
      <genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />
      <headline>Pitcher Preview: Arizona Diamondbacks (29-23) at Philadelphia
        Phillies (26-24), 7:05 p.m.
      </headline>
      <description></description>
    </itemRef>
    <itemRef href="xt.5932656-preview" contentType="application/nitf+xml">
      <genre type="spcpnat:spfixt" qcode="spfixt:pre-event-coverage">
        <name xml:lang="en-US">Game Preview</name>
        <broader qcode="dclass:event-summary" />
      </genre>
      <genre type="spcpnat:dclass" qcode="dclass:event-summary" />
      <genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />
      <headline>Preview: Arizona Diamondbacks (29-23) at Philadelphia
        Phillies (26-24), 7:05 p.m.
      </headline>
      <description role="drol:summary">A pair of teams coming off big weekend
        sweeps will square off tonight at Citizens Bank Park, where the
        Philadelphia Phillies welcome the Arizona Diamondbacks for the start of a
        three-game series.
      </description>
    </itemRef>
  </group>
</groupSet>
</packageItem>

```

18 Exchanging News: News Messages

18.1 Introduction

News communication is often by means of a provider's broadcast or multicast feed, with transactions carried out by a separate sub-system that has its own specific identification and transmission methods, independent of the G2 standards.

The News Message (<newsMessage>) facilitates these transactions, enabling the exchange of any number of NewsML-G2 Items by any kind of digital transmission system.

The use of <newsMessage> is entirely optional: items may also be exchanged using SOAP, Atom Publishing Protocol (APP) or other suitable content syndication method.

News Messages are a convenient wrapper when sending a Package Item and all of its referenced content in a single transaction (but note that it is not necessary to send packages and associated content together in this manner).

The News Message properties are transient; they do not have to be permanently stored, although it may be useful if transmission/reception records are maintained, at least for a limited period during which requests for re-sending of messages may be made. If the content items carried by the message have a dependency on news message information, this information should be stored with the items themselves.

18.2 Structure

A diagram of the structure of a News Message and its payload of Items is shown in **Figure 20** below.

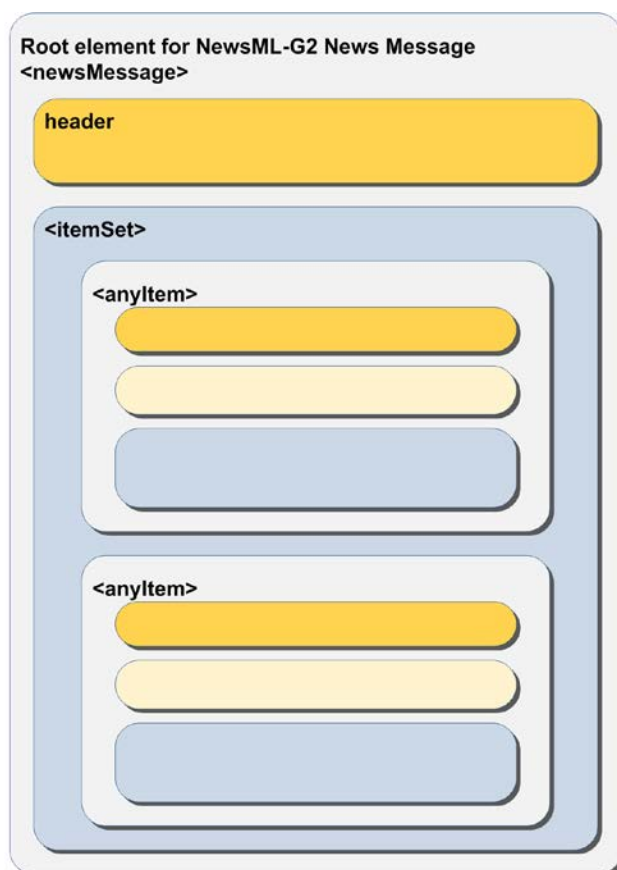


Figure 20: News Messages can convey any number of any kinds of G2 Item

A News Message **MUST** have a <header> component, which **MUST** contain a <sent> transmission timestamp. These are the only mandatory requirements.

The News Message payload is the <itemSet> component, which may have an unbounded number of child <newsItem>, <packageItem>, <conceptItem> and <knowledgeItem> components in any combination and in any order.

18.2.1 News Message Header <header>

The Header contains a set of properties useful for managing the transmission and reception of news. The simple implementation of News Message uses elements with string values, e.g.

```
<sender>thomsonreuters.com</sender>
```

Optionally, providers can use more semantically defined values for header properties, either using URI, QCode or literal values, and also use other QCode type properties to add further detailed information, e.g.:

```
<sender qcode="sdr:rtr">thomsonreuters.com</sender>
```

or

```
<sender literal="reuters">thomsonreuters.com</sender>
```

If using controlled values for properties of the News Message, providers should note the following conditions:

- ❖ If QCode values are used: either a <catalog> or a <catalogRef> child elements of <header> MUST be present to provide the required Scheme URIs of the controlled vocabularies used. (see [Creating and Managing Catalogs](#) for details)¹³. The scope of the <header> catalog is limited to the News Message header only, and is NOT inherited by any of the Items under <itemSet>.
- ❖ If a @literal attribute is used together with a string value for a property: the value of the @literal and of the string should be identical (as in the second example of <sender> above); if not, then the value of @literal takes precedence.

The child elements of <header> that may use qualifying attributes are:

- ❖ sender
- ❖ origin
- ❖ channel
- ❖ destination
- ❖ timestamp (optional @role, see below)
- ❖ signal (may use @qcode only; PCL only)

The following qualifying attributes are defined:

Name	Cardinality	Datatype	Definition
qcode	0..1	QCodeType	A qualified code assigned as a property value.
uri	0..1	XML Schema anyURI	A URI which identifies a concept.
literal	0..1	XML Schema normalizedString	A free-text value assigned as a property value.
type	0..1	QCodeType	The type of the concept assigned as a controlled or an uncontrolled property value.
role	0..1	QCodeType	A refinement of the semantics of the property.

Both string values of properties and the use of qualifying attributes are given as examples below. Note that unless stated otherwise, the scheme aliases used in QCode type properties are fictional.

¹³ <catalog> or <catalogRef> are optional for News Messages, but mandatory for Items.



18.2.1.1 Sent timestamp <sent>

Mandatory, non-repeatable. The transmission timestamp is in ISO8601 format. Note that if a news message is re-transmitted, the sender is NOT required to update this property.

```
<sent>2010- 10- 19T11: 17: 00. 100Z</sent>
```

18.2.1.2 Catalog <catalog>

An optional reference to the Scheme URI of a Controlled Vocabulary to enable the resolution of QCode values of properties used in the document:

```
<catalog>  
  <scheme alias="nmsig" uri="http://cv.example.com/newsmlcodes/newsmsgsignal/" />  
</catalog>
```

18.2.1.3 Remote Catalog <catalogRef>

An optional reference to a remote catalog containing one or more scheme URIs for the CVs used in the document:

```
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
```

18.2.1.4 Message Sender <sender>

An optional, non-repeatable string. Best practice is to identify the sender (not necessarily the same as the provider in <itemMeta>) by domain name

```
<sender>thomsonreuters.com</sender>
```

18.2.1.5 Transmission ID <transmitId>

An optional, non-repeatable string identifying the news message. The ID must be unique for every message on a given calendar day, except that if a message is re-transmitted, it may keep the same ID. The structure of the string is not specified by the IPTC.

```
<transmi tId>tag: reuters. com, 2009: newsml _0VE488500- PKG</transmi tId>
```

18.2.1.6 Priority <priority>

Optional, non-repeatable. The sender's indication of the message's priority, restricted to the values 1-9 (inclusive), where 1 is the highest priority and 9 is the lowest. Priority would be used by a transmission system to determine the order in which a message is transmitted, relative to others in a queue. Urgency is an editorial judgement expressing the relative news value of an item.

```
<pri ori ty>4</pri ori ty>
```

Note that message Priority is not the same as the <urgency> of an individual Item within the news message, although they may be correlated. A typical example is that the Priority may be set to the highest Urgency value associated with any of the Items carried by the message.

18.2.1.7 Message Origin <origin>

An optional, non-repeatable string whose structure is not specified by the IPTC. It could denote the name of a channel, system or service, for example.

```
<ori gi n>MMS_3</ori gi n>
```

or

```
<ori gi n qcode="nmorgn: mms3" />
```

18.2.1.8 Channel Identifier <channel>

An optional, repeatable string that gives the partners in the news exchange the ability to manage the content, for example as part of a permissions or routing scheme. The notion of a channel has its origins in



multiplexed systems, but may be an analogy for a product – a virtual <channel> delivered within a physical stream identified by <destination>.

```
<channel>email</channel>
```

or

```
<channel qcode="nmchan:smtp" />
```

18.2.1.9 Destination <destination>

An optional, repeatable, destination for the news message, using a provider-specific syntax. In broadcast delivery systems, this may be one or more physical delivery points:

```
<destination>foo.bar@example.com</destination>
```

or

```
<destination role="nmdest:foobar">UKI</destination>
```

18.2.1.10 Timestamps <timestamp>

An optional and repeatable <timestamp> property may be used to indicate the date-time that a <newsMessage> was received and/or transmitted. The values must be expressed as a date-time with time zone. The property may be refined using the optional @role, which may be a literal or QCode value, but note that if using a QCode the string that should be interpreted as a QCode MUST be defined outside the NewsML-G2 specification by the provider of the News Message..

```
<timestamp role="received">2010-10-18T11:17:00.000Z</timestamp>
<timestamp role="transmitted">2010-10-19T11:17:00.100Z</timestamp>
```

18.2.1.11 Signal <signal>

Optional, repeatable; at PCL only, the <signal> element can be used to indicate any special handling instruction to the receiver's G2 processor:

```
<signal qcode="nmsig:atomic" />
```

There is a recommended IPTC News Message Signal NewsCode CV with a Scheme URI of:

<http://cv.iptc.org/newscodes/newsmsgsignal/>

The recommended scheme alias is "nmsig" and there is currently one member of the CV: "atomic" which signals to the receiving processor that the content of the News Message should be processed together. This enables providers to indicate that a Package Item and all of the Items that are referenced by that Package should be processed as a single "atomic" (i.e. indivisible) bundle.

18.2.2 News Message payload <itemSet>

Each <newsMessage> MUST have one <itemSet> component, which wraps the G2 Items that are to be transmitted.

The contents of <itemSet> can be "any" component or property from the NAR namespace. This is to enable schema-based validation of the content of the Items in the message.

The IPTC recommends that the child elements of <itemSet> be any number of <newsItem>, <packageItem>, <conceptItem>, and/or <knowledgeItem> components, in any combination, in any order.

The listing below shows a "skeleton" News Message containing a Package Item that references four News Items, together with the referenced items themselves, as an "atomic" package, with some of the <header> properties using a QCode value, which requires a reference to a Catalog (in this case two Remote Catalogs referenced using <catalogRef>).

LISTING 24 News Message conveying a complete News Package

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsMessage xmlns="http://iptc.org/std/nar/2006-10-01/>
  <header>
    <sent>2010-10-19T11:17:00.100Z</sent>
```



```
<catalogRef href="http://www.example.com/std/catalog/NewsMessages_1.xml" />
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_25.xml" />
<sender>thomsonreuters.com</sender>
<transmitId>tag:reuters.com, 2009:newsml_OVE488500-PKG</transmitId>
<priority>4</priority>
<origin>MMS_3</origin>
<destination role="nmdest:foobar">UKI</destination>
<channel>TVS</channel>
<channel>TTT</channel>
<channel>WWW</channel>
<timestamp role="received">2010-10-18T11:17:00.000Z</timestamp>
<timestamp role="transmitted">2010-10-19T11:17:00.100Z</timestamp>
<signal qcode="nmsig:atomic" />
</header>
<itemSet>
  <packageItem>
    <itemRef residref="N1" />
    <itemRef residref="N2" />
    <itemRef residref="N3" />
    <itemRef residref="N4" />
  </packageItem>
  <newsItem guid="N1">.....</newsItem>
  <newsItem guid="N2">.....</newsItem>
  <newsItem guid="N3">.....</newsItem>
  <newsItem guid="N4">.....</newsItem>
</itemSet>
</newsMessage>
```




This page intentionally blank



19 Migrating IPTC 7901 to NewsML-G2

19.1 Introduction

The text transmission standard IPTC 7901 (and ANPA1312, with which it is closely associated) were a mainstay of the exchange of text for 30 years. The following tables map IPTC 7901 fields to their equivalent in either NewsML-G2 News Item or News Message.

In some cases there is a choice of field mappings. For example there is only one timestamp field in IPTC 7901, but several in NewsML-G2.

When providers convert to NewsML-G2, they should determine the source of the metadata used to populate the IPTC 7901 field. For example, if from an editorial system, the field label or usage may indicate the NewsML-G2 property to which the 7901 timestamp should be mapped.

Customers converting IPTC 7901 fields to NewsML-G2 properties should consult their providers' documentation, or make inquiries about the mapping of IPTC 7901 fields in the provider's systems.

19.2 Message Header

Field Name	Example	NewsML-G2 Property and Example
Source Identification	FRS	<p>The 7901 definition is "to identify the news service of the originator and the message." An equivalent NewsML-G2 property would be the <code><destination></code> property of News Message:</p> <pre><destination>FRS</destination></pre> <p>An alternative property would be <code><channel></code></p> <pre><channel>FRS</channel></pre> <p>Within a News Item, there is a <code><service></code> property that may also be used, however, the IPTC recommends that Source Identification expressed a distribution intention, not merely a service description. Therefore its natural home is in News Message.</p>
Message Number	9999	<p>In IPTC 7901, this may be up to four digits and therefore cannot be guaranteed to be unique. Some providers reset the counter to zero every 24 hours. The intention is to be able to identify a message within a sequence of messages.</p> <p>The equivalent G2 field would be <code><transmitId></code> of News Message</p>
Priority of Story	3	<p>IPTC 7901 specifies a single digit of value 1 to 6 inclusive. NewsML-G2 has two equivalents:</p> <pre><urgency></pre> in the Administrative Metadata group of <code><contentMeta></code> <pre><urgency>3</urgency></pre> <pre><priority></pre> in the <code><header></code> of News Message. <pre><priority>3</priority></pre> <p>If the 7901 field is mapped from an editorial system and is under the control of journalists, then <code><urgency></code> would seem to be appropriate, since this expresses a journalistic intention.</p> <p>However, some 7901 priority fields may be set by a transmission system using some other criteria, such as the type of service, and in any case these values may be linked in order that journalists can move urgent stories to the head of a priority queue.</p>
Category of Story	POL	<p>In IPTC7901, this could be up to three characters. A provider converting its categories for NewsML-G2 would create a Controlled Vocabulary mapping</p>



Field Name	Example	NewsML-G2 Property and Example
		<p>the legacy codes to a Knowledge Item – thus allowing the receiver to get more information about the code, or to download all of the codes in order to populate a GUI menu. The CV would be expressed using <subject> in Content Metadata</p> <pre><subject type="cpnat: abstract" qcode="cat: POL"> <name>Politics</name> </subject></pre>
Word Count	195	<p>The maximum value allowed in IPTC 7901 is 9999. Word Count is one of the properties of <i>News Content Characteristics</i> group of properties. (See the NewsML-G2 Specification at www.newsml-g2.org for details.) The plain text of IPTC 7901 is an attribute of the Inline Data component <inlineData></p> <pre><inlineData wordcount="195"> Text text text </inlineData></pre> <p>There is no limit to the number of words that can be expressed in G2</p>
Optional Information	(optional)	<p>These two fields are sometimes used by providers in their own specific format or syntax, so it is not possible to give a definitive mapping from IPTC 7901 to NewsML-G2.</p>
Keyword/Catch-Line		<p>Providers converting from IPTC 7901 to G2 should determine the source of the information that is being put into the 7901 fields, and map to the appropriate NewsML-G2 property.</p> <p>Customers converting an IPTC 5901 service into NewsML-G2 and who are unsure about which of NewsML-G2's properties to populate with the 7901 information should consult the provider's documentation, or inquire of the provider,</p> <p>Part of the content of the Keyword/Catch-Line field may be equivalent to the NewsML-G2 <slugline> property in the Descriptive Metadata group of <contentMeta>.</p> <pre><slugline separator="-">US- POLI TICS/BUSH</slugline></pre>

19.3 Message Text

If the message text is to be conveyed as plain text, then the <inlineData> wrapper would be used. If the migration to NewsML-G2 also involves migration of the text to some mark-up such as NITF or XHTML, this would be conveyed in the <inlineXML> wrapper.

19.4 Post-Text Information

The IPTC7901 post-text fields all relate to date-time. As previously recommended, providers should examine the origin of the information being used to fill the IPTC 7901 day and time field, and use the appropriate NewsML-G2 field for that information. The following are possibilities:

- ❖ If Day and Time is populated by the transmission system, the <sent> property of News Message would be an appropriate mapping.
- ❖ If the IPTC 7901 field represents the only timestamp information available from the provider's system, then the provider MUST put this information in the <versionCreated> timestamp property of <itemMeta> since its use is a mandatory requirement of NewsML-G2,
- ❖ If versioning of Items is being implemented, then the original timestamp may be preserved in <firstCreated>

These properties are of XML date time type so must have the FULL date and a time expressed, with time zone:

```
<versionCreated>2009-02-09T12:30:00Z</versionCreated>
```



If the provider believes that the IPTC 7901 day and time is being populated from the timestamp on the story itself, then subject to the rule of <versionCreated> being followed, the appropriate properties would be <contentCreated> or <contentModified>.in the Management Metadata group of <contentMeta>.

If versioning is being implemented, the original day and time may be preserved in <contentCreated>, If <contentModified> is used, then <contentCreated> SHOULD also be present

Both properties use TruncatedDateTime property type, allowing the value to be truncated from the right.

| <contentCreated>2009- 02- 09</contentCreated>

Field Name	Example	NewsML-G2 Property and Example
Day and Time	091230	Six digits are be used to express the day and time in IPTC 7901, in practice this means a format of DDHHMM, where DD is the day of the month, HH is the hour, MM minutes. In G2, there are more timestamp fields available,
Time Zone	GMT	Three-alpha character field. Optional in IPTC 7901. No separate equivalent in NewsML-G2 (none needed)
Month of Transmission	Feb	Three-alpha character field. Optional in IPTC 7901. No separate equivalent in NewsML-G2 (none needed)
Year of Transmission	09	Two-digit field. Optional in IPTC 7901. No separate equivalent in NewsML-G2 (none needed)



This page intentionally blank

20 Designing a NewsML-G2 feed from scratch

20.1 Introduction

When designing a NewsML-G2 feed, some fundamental questions need to be asked, including:

- ❖ What is the provider's business value proposition and how must this be supported by the NewsML-G2 implementation?
- ❖ What are the needs and capabilities of customers? Sometimes, it is the customers who are driving the move the G2; if so, happy days! But more often, customers will have some issues and will need sympathetic handling.
- ❖ What limitations – if any – do the provider's existing systems and processes impose on the NewsML-G2 project?

This chapter gives general guidance on the issues an implementer of a feed might encounter, and where necessary points to more detailed information on specific topics described elsewhere in the Guidelines.

As a general principle, the IPTC recommends that providers strive for inter-operability, so where possible, stick to using the recommended “mainstream” properties as identified by the [Property Usage](#) sheet below. The sheet lists the NewsML-G2 properties which are most widely used by currently existing NewsML-G2 feeds. As there may be more than one way to implement a feature; in such cases the IPTC recommends choosing the most widely-used property as indicated by the sheet.

Also consider using existing IPTC NewsCode Schemes for property values where appropriate, in preference to creating proprietary schemes that accomplish a similar purpose. A table showing [IPTC Controlled Vocabulary Usage](#) is shown below.

20.2 Content-driven design

The design of a NewsML-G2 feed will depend on the content type being conveyed, since each type will have specific metadata and processing needs.

For all content types, note the basic NewsML-G2 rule: a single logical item of content per News Item, which can be represented by more than one technical representation. For multiple pieces of content in a named relationship, use the Package Item (see [Chapter 8](#)) and for sending one or more G2 Items in a transmission workflow, the [News Message](#) may be used.

20.2.1 Content choices

20.2.1.1 [Text](#)

The most commonly-used NewsML-G2 container for text is <inlineXML> wrapping XHTML content, although NITF is also used. Note that any well-formed XML may be conveyed in <inlineXML> so this can be aligned to a more specific business purpose such as XBRL (Extended Business Reporting Language).

It is also possible to convey a text-based object as binary content, for example as a PDF, using <remoteContent> or as plain text using <inlineData>.

When converting from legacy text formats such as IPTC7901, the initial advantages of NewsML-G2 include the ability to separate metadata from the content, and the use of globally-unique controlled values to classify content.

Going further, there are more specialised structures available for implementers to provide financial information, express quantitative data, and give precise geographic details in a machine-readable form.

20.2.1.2 [Pictures](#)

As binary content, pictures (and graphics) are conveyed using the <remoteContent> wrapper. The [Quick Start: Pictures and Graphics](#) chapter gives an overview of NewsML-G2 features and the relationship between G2 metadata and embedded metadata such as the “IPTC Core” fields in JPEG/TIFF files. There is also a chapter on Mapping Embedded Photo Metadata to NewsML-G2 that provides more detailed information on this issue.



20.2.1.3 Video

The first consideration for video is the business requirement, since this drives the technical design in one of two directions:

- ❖ Simple one-part video – probably delivered in multiple formats to suit different technical platforms
- ❖ Multi-part video delivered as a reference to a single file (and potentially multiple technical renditions) that requires each part or shot to have its own administrative and descriptive metadata.

NewsML-G2 serves both of these content models, which are described in the [Quick Start: Video](#) chapter.

20.3 The Basics

The Chapter on [NewsML-G2 Basics](#) has a description of the mandatory and commonly-used properties.

20.4 Identification and version control

Choose a scheme for generating the @guid of G2 Items. The IPTC has registered a namespace urn “newsml” and a specification for this purpose (see [Item Identifier](#)) but other schemes such as TAG URI are available.

Decide how Item version will be generated: it must start at “1” and ideally should increment by “1” each time a new version is published, although a new version may have any number that is greater than the previous version number. Providers may also use <link> to inform receivers of the previous version and provide a location or id for it (see [Using Links](#) for more details on this)

20.5 Conformance

The Core Conformance Level (CCL) is a minimal sub-set of G2 that can usefully get work done, and is the default if @conformance is omitted. In practice, most implementers use the properties of “Power” conformance (PCL) as it is much more powerful in expressing metadata and helps to implement the requirements of the Semantic Web. Ensure that if using PCL, the @conformance is present and set to “power” in all G2 Items, otherwise end-users may not see the results you expect as in these circumstances a G2 processor should default to CCL and ignore PCL properties! There is no rule for this enforced by the schema files, so this situation will not currently be caught by validation during testing of the feed.

20.6 Validation

When designing and testing the feed, validation against the IPTC scheme files will be essential. The IPTC recommends that validation should not be required in production: it can be resource intensive and introduces a potential point of failure. If a provider chooses to validate production documents, they must NOT validate against the schema files hosted by the IPTC for production documents.

20.7 Catalogs and Controlled Vocabularies.

Even after following the advice about using NewsCodes (above) it is almost certain that some provider specific schemes will be needed. See the section on [Controlled Vocabularies](#) for detailed information on how to create and maintain CVs and Catalogs.

20.8 Rights Information

See [Chapter 23 Rights Metadata](#) to go beyond the basics of expressing rights information.

20.9 Management Metadata

20.9.1 Timestamps

There are four timestamps available in NewsML-G2 documents, two are properties of Item Metadata and relate to management of the Item; two are properties of Content Metadata and are part of the Administrative properties of the Item’s content. The four properties are:

- ❖ itemMeta/firstCreated
- ❖ itemMeta/versionCreated



- ❖ contentMeta/contentCreated
- ❖ contentMeta/contentModified

NewsML-G2 timestamps in Item Metadata express the full date and time of the creation or update of the Item. Content Metadata timestamps may be truncated from the right because the precise date and time of the creation of content (for example an historic photograph) may be unknown. Note that any timestamp that contains the time **MUST** contain a time offset or Z for UTC, for example the format YYYY-MM-DDThh:mm:ss(+/-)hh:mm:ss, or YYYY-MM-DDThh:mm:ssZ.

The single mandatory timestamp for a G2 document is <versionCreated>, but the following rules **MUST** be followed when any other timestamp is present.

20.9.1.1 Data Value rules

1. If <contentCreated> is used, it **MUST NOT** be later than <versionCreated>.
2. If <contentModified> is present, <contentCreated> **SHOULD** also be present and **MUST** be earlier than <contentModified>.
3. If <contentModified> is present, it **MUST NOT** be later than <versionCreated>.
4. If <firstCreated> is present, it **MUST NOT** be later than <versionCreated>.

20.9.1.2 Data Processing rules

1. The recipient processor **MUST** first check if a <contentModified> element is present.
2. If not it **MUST** check if a <contentCreated> element is present.
3. If not it **SHOULD** assume that the content was created at the time indicated by <versionCreated> element in itemMeta.

For a full description of the timestamps and processing rules, see the NewsML-G2 Specification, which may be downloaded by visiting www.newsml-g2.org and navigating to Specification resources..

A number of existing implementations use only itemMeta/versionCreated for text, but for picture, graphics and video, it is generally accepted that contentMeta/contentCreated and contentMeta/contentModified should be used as, for example, the moment when a picture is taken is considered significant.

20.9.2 Publishing Status/Embargoed

Both properties of <itemMeta> have “hidden” values, in that there are default values that are assumed if the actual properties are absent. Publishing Status defaults to “usable” and Embargoed to “none”. There are separate topics for these properties in [Chapter 26 Generic Processes and Conventions](#).

20.9.3 Links

Links identify and can also locate G2 Items and other non-G2 resources that may be used to supplement the understanding of an Item’s content. The default relationship to the target Item is “See Also” but other relationships such as “Translated From” and “Derived From” are defined in the IPTC Item Relationship NewsCodes.

20.10 Property Usage

The following property usage table was compiled from a survey of IPTC members.

Property name	Text	Photo / Graphic	Video	News Package
News Message header	High	High	High (in package)	Medium
sent	Mandatory	Mandatory	Mandatory	Mandatory
sender	High	Medium+	High	Medium
transmitId	Medium	Medium+	High	High
priority	High	Medium	Medium	Medium

Property name	Text	Photo / Graphic	Video	News Package
origin	Medium	Medium	No	Medium
timestamp	Medium	Low+	Medium	No
destination	High - with channel	Medium+ - with channel	Medium	Medium
channel	High - with destination	Medium+ - with destination	High	Medium
News Item				
@guid	Mandatory	Mandatory	Mandatory	Mandatory
@version	Always	Always	Always	Always
catalogRef	Mandatory	Mandatory	Mandatory	Mandatory
catalog	Mandatory	Mandatory	Mandatory	Mandatory
hopHistory	Low	Low	No	No
rightsInfo				
accountable	Low	Low	No	Medium
copyrightHolder	High	High	High	High
copyrightNotice	High	High	High	Medium
usageTerms	Medium	Low+	Medium	Low+
itemMeta				
itemClass	Mandatory	Mandatory	Mandatory	Mandatory
provider	Mandatory	Mandatory	Mandatory	Mandatory
versionCreated	Mandatory	Mandatory	Mandatory	Mandatory
firstCreated	Medium	High	High	Medium+
embargoed	Always	Always	Always	Always
pubStatus	High	High	High	High
role	Medium	Low+	High	Medium
fileName	Medium	Medium	Medium	Medium
generator	Medium	Medium	High	High
profile	Medium	Low+	High	Medium+
service	Medium	Low+	High	Medium
title	Medium	Low	No	Medium
edNote	High	Low+	Low+	No
memberOf	Medium	Low	No	No
instanceOf	Medium	Low	No	No
signal	High	Medium+	High	Medium
altRep	No	No	No	No
deliverableOf	No	No	No	No
hash	No	No	No	No
link	Medium	Low	Medium	Low
contentMeta				
icon	No	Low	Low	No
urgency	High	Medium+	High	Medium
contentCreated	Medium	Medium+	Medium	Medium
contentModified	Medium	Medium	Low	Low+
located	High	Low	High	Low+

Property name	Text	Photo / Graphic	Video	News Package
infoSource	High	High	High	Medium
creator	High	High	Medium	Medium
contributor	Medium	Medium+	Medium	Low+
audience	Medium	No	No	Low+
exclAudience	Low	Low	No	No
altId	High	High	High	Medium+
rating	No	Low	No	No
userInteraction	No	No	No	No
language	Medium	Medium	High	Medium+
genre	Medium+	Low	Medium	No
keyword	Medium+	Low	No	Medium
subject	High	High	High	High
bag	No	No	No	No
slugline	High	Medium+	High	High
headline	High	High	High	High
dateline	Medium+	Low	Low	Medium+
by	Medium	Medium+	Medium	No
creditline	Medium	High	No	Medium
description	Medium	High	High	Medium
partMeta	No	Low	Low	No
assert	~	~	~	~
inlineRef	~	~	~	~
derivedFrom	Low	No	No	No
contentSet				
inlineXML	High	n/a	n/a	n/a
inlineData	Low	n/a	n/a	n/a
remoteContent	No	High	n/a	n/a
groupSet				
main/root Group	n/a	n/a	High	High
subgroups	n/a	n/a	High	High
itemRef	n/a	n/a	High	High

Key	
Mandatory	Mandatory per the NewsML-G2 specification
Always	Explicitly defined, or implied by absence
High	Expected to be present
Medium	May be present
Low	Unlikely to be present (currently)
No	Not present (currently)
~	Technical helpers: These provide richer information about another explicit property
n/a	Not applicable
red	Mandatory mark-up



20.11 IPTC Controlled Vocabulary Usage

The mark-up in **red** indicates mandatory usage

Property name	IPTC CVs (alias)	Provider CVs
News Message header		
destination	-	Medium
News Item		
catalog info	-	-
rightsInfo	-	No
itemMeta		
itemClass	(ninat)	n/a
provider	(nprov)	No
pubStatus	(stat)	n/a
role	No	Medium
generator	-	Medium
service	-	High
edNote	-	Medium
memberOf	-	High
instanceOf	-	High
signal	(sig)	High
link	(irel)	High
contentMeta		
located	-	High
infoSource	(isrol)	Medium
creator	-	High
contributor	-	High
audience	-	Medium
exclAudience	-	High
altId	-	Medium
genre	No	High
subject	(medtop)	-
	(subj)	
	(cpnat)	
	(howextr)	
	(why)	-
description	(drol)	Medium
partMeta	-	Medium
(tech helpers)		
assert	-	High
contentSet		
remoteContent	(rnd)	High
groupSet		
itemRef	-	High

Key	
(alias)	Mandatory per specification
High	Expected to be present
Medium	May be present
No	Available but not used
~	Unavailable
n/a	Not applicable
(alias)	Alias of IPTC recommended scheme
red	Mandatory mark-up

21 News Industry Text Format (NITF)

21.1 Introduction¹⁴

NITF is an XML-based standard for marking up text articles that includes the ability both to create a structure for the content, and to provide a clear separation of content and metadata.

NITF differs from NewsML-G2:

- ❖ NITF provides a mark-up structure for text; NewsML-G2 is a container for news, is content-agnostic and has no mark-up for text content.
- ❖ NITF expresses a single piece of content and does not support, for example, alternative renditions or different parts of the same content.
- ❖ NITF properties have tightly-defined meanings; NewsML-G2 makes more use of generic structures that can be adapted for different uses. Therefore there may not be a direct equivalent in NewsML-G2 for a given NITF property, but a generic structure may be used to express the provider's intention equally effectively.

The following table summarises the major differences between the two standards.

Feature	NITF	NewsML-G2
Content mark-up	✓	✗
Content structure	✓	✗
Extensible	✗	✓
Any content	✗	✓
Management metadata	✓	✓ +
Administrative metadata	✓	✓ +
Descriptive metadata	✓	✓ +
Packages	✗	✓
Alternative renditions	✗	✓
Concepts/Relationships	✗	✓

NITF is a popular and valid choice for the management and mark-up of text in a news environment. This Chapter assumes that organisations wish to continue to use NITF for the mark-up of text, but to migrate to NewsML-G2 as a management container. There are three basic migration paths:

- ❖ Minimum migration: retain NITF and all of its structures in place, and use a minimal NewsML-G2 News Item as a container for NITF documents. NewsML-G2 would convey the complete original NITF document.
- ❖ Partial migration: copy or move some of NITF's metadata structures to the equivalent NewsML-G2 properties, retaining some or all NITF metadata structures and structural mark-up of the text content.
- ❖ Complete migration: move the NITF metadata to the NewsML-G2 equivalents, using NewsML-G2 to convey a text document with NITF content mark-up.

For a minimum migration, implementers need only to refer to the Quick Start Guides to [NewsML-G2 Basics and Text](#).

¹⁴ This Guide refers to NITF v 3.5. A new version 3.6 was released in January 2012. Please see www.nitf.org for further details of NITF and to download a package of the new version and its documentation



For partial or complete migration, the starting point is a consideration of how the structure of NITF relates to NewsML-G2, to form a high-level view of which NITF properties are to be migrated, and where they fit in NewsML-G2 framework.

21.2 Overview of NITF and NewsML-G2 equivalents

NITF documents are composed of the following components:

- ❖ The root element: <nitf>
- ❖ <head>, a container for metadata about the document as a whole
- ❖ <body> a container for content and metadata which is intended for display to the user. The <body> component is further split into:
 - ❖ <body-head> Metadata about the content
 - ❖ <body-content> The content itself
 - ❖ <body-end> Information that is intended to appear at the end of the article.

21.2.1 Root element <nitf>

The root element may uniquely identify the document using @uno, which is equivalent to NewsML-G2 @guid.

21.2.2 Document metadata <head>

<head> may have a @id attribute, a local identifier for the element. Most of the child elements of <head> have equivalents in the NewsML-G2 <itemMeta> component, with exceptions noted in the following table:

NITF element	NewsML-G2 equivalent	Comment
<title>	<title>	The properties are equivalent in NITF and NewsML-G2
<meta>	<itemClass>	<itemClass> uses an IPTC NewsCode to identify the type of content being conveyed.
<toobject>	<subject>	This is part of NewsML-G2 <contentMeta>
<iim>	None	Deprecated in NITF, no NewsML-G2 equivalent
<docdata>	Various	A container for detailed metadata for the document. contains child elements as tabled in 21.2.2.1 below
<pubdata>	None	NewsML-G2 contains no direct equivalent of Publication Metadata. In NITF, the property has many attributes intended to convey information about print publications
<revision-history>		Attributes can express the date and reason for the revision, the name and job title of the person who made the revision

21.2.2.1 <docdata> child elements

These convey detailed management information about the document.

NITF element	NewsML-G2 equivalent	Comment
<correction>	<signal>	Indicates that the item is a correction, identifies the previously published document and any correction instruction. NewsML-G2 <signal> contains a QCode so that processing instructions are constrained by a controlled vocabulary. The property may have a <name> child element and be accompanied by <edNote>



NITF element	NewsML-G2 equivalent	Comment
		conveying a human-readable message about the instruction.
<ev-loc>	<subject>	<ev-loc> contains information about where the news event took place. The NewsML-G2 equivalent of <subject> may use a @role to indicate that the subject is a geographic place. If used, the IPTC Concept Type NewsCode would be either "geoArea" or "poi" (Point of Interest)
<doc-id >	G2 Item @guid	Unique persistent identifier for the document
<del-list >	<timestamp>	Audit trail of points involved in delivery of the document. May be expressed in the <timestamp> property of a NewsML-G2 News Message <header>.
< urgency>	< urgency>	Editorial urgency of the document. NITF uses values 1-8, NewsML-G2 uses 1-9. The value may also be linked to Priority <priority> in a NewsML-G2 News Message.
< fixture>	< instanceOf>	Regularly occurring and/or frequently updated content, for example "Noon Weather Report" or "Closing Stock Market Prices". In NewsML-G2, this is a Flex1PropType property type (PCL only), a template that allows the use of a controlled or uncontrolled value.
< date.issue>	Various	The default NewsML-G2 property would be <versionCreated>, but depending on the provider's intention, <firstCreated>, <contentCreated> or <contentModified> could be used.
< date.release>	Various	NewsML-G2 does not have a direct equivalent for this property. In NITF, the default value is the time of receipt of the document. Where a provider wishes to specifically delay a release, the NewsML-G2 <embargoed> property may be used. If the provider wishes to express a timestamp of the time that the document was released, a timestamp property such as <sent> in News Message should be used.
<date.expire >	@validto	At PCL, the <remoteContent> wrapper may have a Time Validity Attribute of @validto.
<doc-scope >	<audience>	Area of interest that is not a Category, for example geographic: "of interest to people living in Toronto". This is not the same intention as a <subject> of Toronto, which means "about" or "linked to" Toronto. If content should be seen by "Equity Traders" but is not of the Subject "Equities", then the NewsML-G2 <audience> property may be used.
<series >	<memberOf>	In NITF, this is intended to indicate that the article is one of a series of articles, its place in the order of the series, out of a total number of articles in the series. The NewsML-G2 equivalent <memberOf> is a Flex1PropType (PCL only) which may take a controlled or uncontrolled value.



NITF element	NewsML-G2 equivalent	Comment
<ed-msg >	<edNote >	As with NITF, the type of message may be indicated by @role in NewsML-G2
< du-key>	None	NITF uses this property for maintaining a network of relationships to other versions of the same article. NewsML-G2 uses GUID, versioning and, if necessary, Links, to maintain this information
<doc.copyright >	< rightsInfo>	In NewsML-G2, the <rightsInfo> container may hold information about the copyright holder, copyright statement, the entity legally accountable for the content, and usage terms.
<doc.rights >	<rightsInfo >	
<key-list >	<keyword>	Each keyword is conveyed in a separate NewsML-G2 <keyword> element.
< identified-content>	< inlineRef>	<p>This container is used in NITF to hold information about entities such as people, organisations, places or events with an @id enabling it to be referenced from within the content mark-up.</p> <p>The NewsML-G2 equivalent <inlineRef> enables a Concept to be referenced by @idrefs, and contains a QCode identifying the Concept in a controlled vocabulary. Locally, any mark-up element (e..g in XHTML capable of using @id may be linked via <inlineRef> to the concept.</p>

21.3 Content metadata <body.head>

The contents of <body.head> a broadly equivalent to those of <contentMeta>, although this is not universally true. The following table lists the child elements of <body.head> and their NewsML-G2 equivalents.

When conveying NITF in a NewsML-G2 document, some properties of <body.head> may be retained in the NITF document, since they may be part of the structural content of the article, such as <headline> (note spelling) and <byline>. These properties may optionally be copied into their NewsML-G2 equivalents.

NITF element	NewsML-G2 equivalent	Comment
<headline>	<headline >	NewsML-G2 <headline> is contained in <contentMeta> and may appear any number of times. It can be refined by using @role to express the NITF notions of main headline <hl1> and subhead <hl2>
<note >	Various	<p>An advisory for editors may be equivalent to NewsML-G2 <edNote>in <itemMeta>, but NITF <note> has a number of uses, including expressing copyright, and may also be publishable.</p> <p>Provider's migrating from NITF to NewsML-G2 will need to identify the intention of their use of <note> and choose from several NewsML-G2 properties that express this intention, for example <rightsInfo> (a property of an Item root element), or <description> (part of <contentMeta>) with a @role defining its purpose.</p> <p>Customers converting an NITF service to NewsML-G2</p>



NITF element	NewsML-G2 equivalent	Comment
		will need to consult with their provider, either directly or via documentation.
<rights >	< rightsInfo>	See 21.2.2.1 above
<byline >	<by>	If using every available feature of NITF <byline>, NewsML-G2 <by> should be implemented at PCL. It is part of <contentMeta>
<distributor >	<provider >	<provider> is a mandatory property of NewsML-G2 <itemMeta>
<dateline >	< dateline>	NewsML-G2 <dateline> is part of <contentMeta>
< abstract>	< description>	NewsML-G2 <description> is part of <contentMeta> and may be refined by @role. The IPTC Description Role NewsCode value would be "summary"
<series >	<memberOf>	See 21.2.2.1 above

21.4 Body Content <body. content>

NewsML-G2 has no specific mark-up for text, and NITF is a valid choice for marking up text to be conveyed by NewsML-G2 <inlineXML>.

LISTING 25 An NITF marked-up article conveyed in <inlineXML>

```
<contentSet>
  <inlineXML contentType="application/nitf+xml">
    <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/"
      xsi:schemaLocation="http://iptc.org/std/NITF/2006-10-18/
        XSD/NITF/3.5/specification/nitf-3-5.xsd">
      <body>
        <body.head>
          <headline>
            <hl1>Arizona Diamondbacks (29-23) at Philadelphia Phillies
              (26-24), 7:05p.m.</hl1>
          </headline>
          <byline>
            <byt1>Sports Network</byt1>
          </byline>
          <abstract>
            <p>A pair of teams coming off big weekend sweeps will square off
              Tonight at Citizens Bank Park, where the Philadelphia Phillies
              welcome the Arizona Diamondbacks for the start of a three-game
              series.</p>
          </abstract>
        </body.head>
        <body.content>
          <p>(Sports Network) - With the Philadelphia Phillies picking
            up momentum after their three-game sweep of the Atlanta Braves,
            and the Arizona Diamondbacks capturing all four of their games
            against the Houston Astros, tonight's game at Citizen Bank Park
            looks set to be a clash of the Titans.</p>
        </body.content>
      </body>
    </nitf>
  </inlineXML>
</contentSet>
```

21.5 End of Body <body.end>

There is no formal equivalent to the <body.end> container in NewsML-G2. It's placement at the end of <body.content> expresses its significance in the format of an article. So it is possible that they would



need to be retained in the NITF content conveyed in <inlineXML>, or expressed in another mark-up language.

The NewsML-G2 equivalent of <tagline> would be <by> with a @role if necessary to refine its use. <bibliography> may be expressed using NewsML-G2 <description> with a @role. Their placement relative to the parent text is a processing issue – NewsML-G2 is format-agnostic.

22 Mapping Embedded Photo Metadata to NewsML-G2

22.1 Introduction

Embedded metadata in JPEG and other file formats have been a *de facto* standard since the 1990s. In fact, the metadata schema adopted by Adobe Systems Inc for the original “File Info” dialog box in Photoshop is based on the IPTC Information Interchange Model (IIM). This standard for exchanging all types of news assets and metadata pre-dates XML-based standards such as NewsML-G2. The IIM-based embedded properties in images became known as the “IPTC Fields” or “IPTC Header” and they were widely adopted in professional workflows.

In about 2001, in order to overcome some technical limitations imposed by this legacy model, Adobe introduced the Extensible Metadata Platform, or XMP, to its suite of applications that includes Photoshop. Adobe also worked with the IPTC to migrate the properties of the legacy “IPTC Header” to XMP. Most of the original IIM-based metadata properties are now contained in the IPTC Core Schema for XMP. Further IPTC properties that are derived from NewsML-G2 properties are available in the IPTC Extension schema for XMP.

Although developed by Adobe, XMP is an open technology like its IIM-based predecessor, and has been adopted by other software vendors and manufacturers. Adobe and others, including Microsoft, Apple and Canon, have formed the Metadata Working Group, a consortium that promotes preservation and interoperability of digital image metadata. The MWG web site is www.metadatatworkinggroup.org.

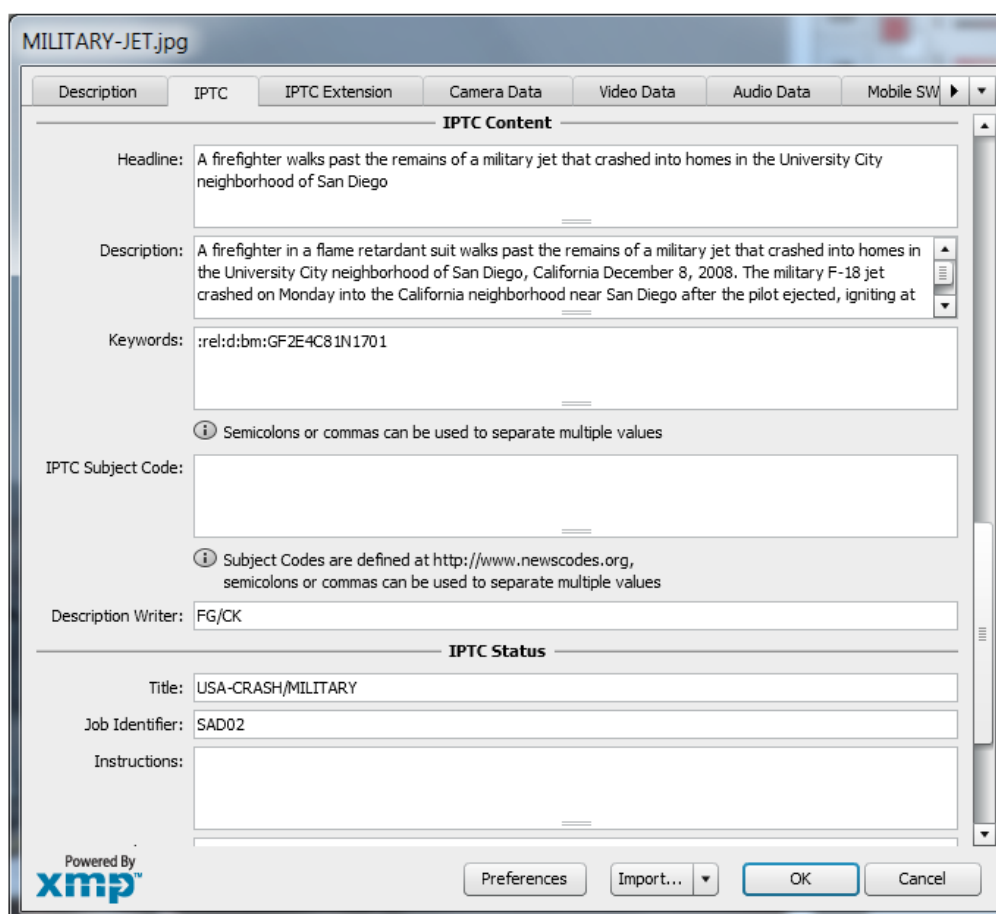


Figure 21: The IPTC Core metadata panel for an image opened in Adobe Photoshop.

22.2 IPTC Metadata for XMP

To use the IPTC Core and Extension metadata in Adobe CS3 and CS4 a set of custom panels for Bridge must be downloaded from the IPTC web site and installed on your Windows or Mac OS X computer. Adobe CS5 and higher provides built-in File Info panels covering both IPTC Core and Extension. Be aware

that these custom panels only work up to version CS 5.5 of Bridge. To download visit www.iptc.org and navigate to Photo Metadata.

The IPTC web site also lists image manipulation software packages that synchronize the legacy IIM-based header fields and XMP properties. The IPTC Custom Panels User Guide tabulates the field labels used by IIM, IPTC Core, and the equivalent fields of software packages such as Photoshop, iViewMedia, Thumbs, and Irfanview.

22.3 Synchronizing XMP and legacy metadata

When opening an image that contains the legacy IIM-based IPTC metadata, using an application that supports XMP, users may need to be aware of synchronization issues. Digital images that pre-date XMP contain the legacy IIM-based metadata in a special area of the JPEG, TIFF or PSD file called the Image Resource Block (IRB). XMP-aware applications store metadata in the file differently, in the XMP Packet.

Current versions of Adobe's software such as Photoshop write the metadata to the IIM-header wrapped by the IRB, and to the XMP packet, when saving a file. Therefore, an image which contained only legacy metadata when opened could have metadata written to a new XMP packet (where an IIM→XMP mapping exists) when saved.

If a picture with XMP metadata is modified by a non-compliant application (i.e. the IRB only is changed), this should be detected by the next XMP-compliant processor that opens the picture, using a checksum. The Metadata Working Group guidance document has a detailed chapter on these synchronization issues.



There is no guarantee that the synchronization of legacy and XMP metadata will be supported in future versions of Adobe software, or applications from other vendors.

22.4 Picture services using IIM/XMP

Many news and picture agencies deliver images to their customers using IIM. Customers receive these services in three variants:

- ❖ As binary files such as JPEG that support embedded IIM metadata in the file header as "IPTC Fields" and/or XMP.
- ❖ As binary files such as JPEG with additional embedded IIM fields beyond the set adopted by Adobe that are used by proprietary image-management software applications. These fields are not read by off-the-shelf software packages such as Photoshop. This format is used by the photo departments of many news agencies.
- ❖ As binary IIM files which consist of the IIM envelope and associated object (picture) data. The conveyed picture may also contain embedded IPTC-IIM fields, but this metadata may not be synchronized with the IIM envelope in all instances. Synchronization practice varies between providers.

22.5 Rationale for moving to G2

It is probable that providers will continue to embed metadata in image and graphics files, but both providers and customers may derive benefits from exchanging this content using NewsML-G2:

- ❖ metadata carried in a NewsML-G2 document is accessible without the need to retrieve and open the associated image file;
- ❖ in some business applications, editors have to modify picture metadata but have no access or permission to modify the metadata in the image file, and the provider needs therefore to instruct receivers to use the NewsML-G2 metadata;
- ❖ partners in an information exchange can standardise on a common method for managing all kinds of news objects.
- ❖ Some news providers' picture services are already being migrated to NewsML-G2. Some are already using NewsML-G2 as their internal standard for storing and exchanging all types of news objects; it makes sense to want to migrate to a common standard for customer-facing applications too.



- ❖ IIM is a legacy format and has some inherent issues, such as limits on field lengths and difficulty with internationalization.

22.6 IIM resources

The IPTC web site has an IIM Home page (visit www.iptc.org and navigate to IIM) which contains links to the Specification and other documents. These include a list of software packages that support IIM metadata fields, and a link to a document that maps IIM fields → IPTC Core (XMP) → Software Package labels, maintained by David Riecks at www.controlledvocabulary.com.

More about the use of IIM for photos can be found in the “Photo Metadata” section of the IPTC website.

22.7 Approach

Although IIM can handle any type of media object, its use for media types other than pictures is now rare, so this section will focus on the issue of migrating IIM picture metadata, chiefly those found in Adobe’s Photoshop “IPTC Header”, to NewsML-G2 properties. It is NOT intended to describe a mapping of the complete IIM envelope to NewsML-G2.

The mapping of IPTC Core Schema properties to NewsML-G2 is documented in the IPTC Photo Metadata Specification 2009 in the Photo Metadata section of the IPTC web site.

22.8 IIM to NewsML-G2 Field Mapping Reference

IIM is organised into Records and DataSets. The DataSets that are embedded in image files are in the Application Record (Record Two). Each IIM DataSet is labelled according to the parent Record and its position within the Record (these numbers are not necessarily contiguous). The IIM Specification contains details of the record structure. (Location details above in paragraph 22.6)

The table below shows IIM DataSets that have been (with noted exceptions) adopted for the IPTC Core (XMP) Photo Metadata. Each DataSet is shown with its IIM name, equivalent IPTC XMP Core Name (if available), and the corresponding G2 Property. Brief notes alongside each property are expanded in describing the implantation of an example picture in 22.9.

DataSet	IIM Name	IPTC Core Name (XMP)	NewsML-G2 Property	Note
2:05	Object Name	Title or Document Title (XMP)	itemMeta/title	
2:10	Urgency	Deprecated	contentMeta/urgency	The original IIM properties have no equivalent in IPTC Core for XMP
2:15	Category	Deprecated	contentMeta/subject	
2:20	Supplemental Category	Deprecated	contentMeta/subject	
2:25	Keywords	Subject	contentMeta/keyword	New in NewsMLG2 2.4
2:40	Special Instruction	Instructions	itemMeta/edNote	Alternative: rightsInfo/usageTerms
2:55	Date Created	Date Created	contentMeta/contentCreated	
2:60	Time Created	Date Created	contentMeta/contentCreated	If present, merge with Date Created
2:80	By-Line	Creator	contentMeta/creator/name	
2:85	By-line Title	Creator's Jobtitle	contentMeta/creator/related	
2:90	City	City	contentMeta/subject	May be used with <assert> (see below)
2:92	Sub-location	Location		
2:95	Province/State	State		
2:100	Country/Primary Location Code	Country Code		
2:101	Country/Primary Location Name	Country		
2:103	Original Transmission Reference	Transmission Reference	itemMeta/altId	If the reference is a Job ID use itemMeta/memberOf
2:105	Headline	Headline	contentMeta/headline	

DataSet	IIM Name	IPTC Core Name (XMP)	NewsML-G2 Property	Note
2:110	Credit	Credit	contentMeta/creditline	
2:115	Source	Source	rightsInfo/copyrightHolder	
2:116	Copyright Notice	Copyright Notice	rightsInfo/copyrightNotice	
2:120	Caption/Abstract	Description	contentMeta/description	
2:122	Writer/Editor	Caption Writer	contentMeta/contributor	A @role should be added to contributor

22.9 IIM to NewsML-G2 Mapping Example

The screenshot below shows fields of the IPTC Panel in Adobe Photoshop's File Info dialog, containing metadata from the IPTC Core metadata schema. Note the neighbouring tab labelled IPTC Extension that contains metadata fields of the IPTC Extension schema. In the example, the IIM DataSets to be mapped to NewsML-G2, and their values, are tabulated, and in the succeeding paragraphs, the mapping discussed in more detail.

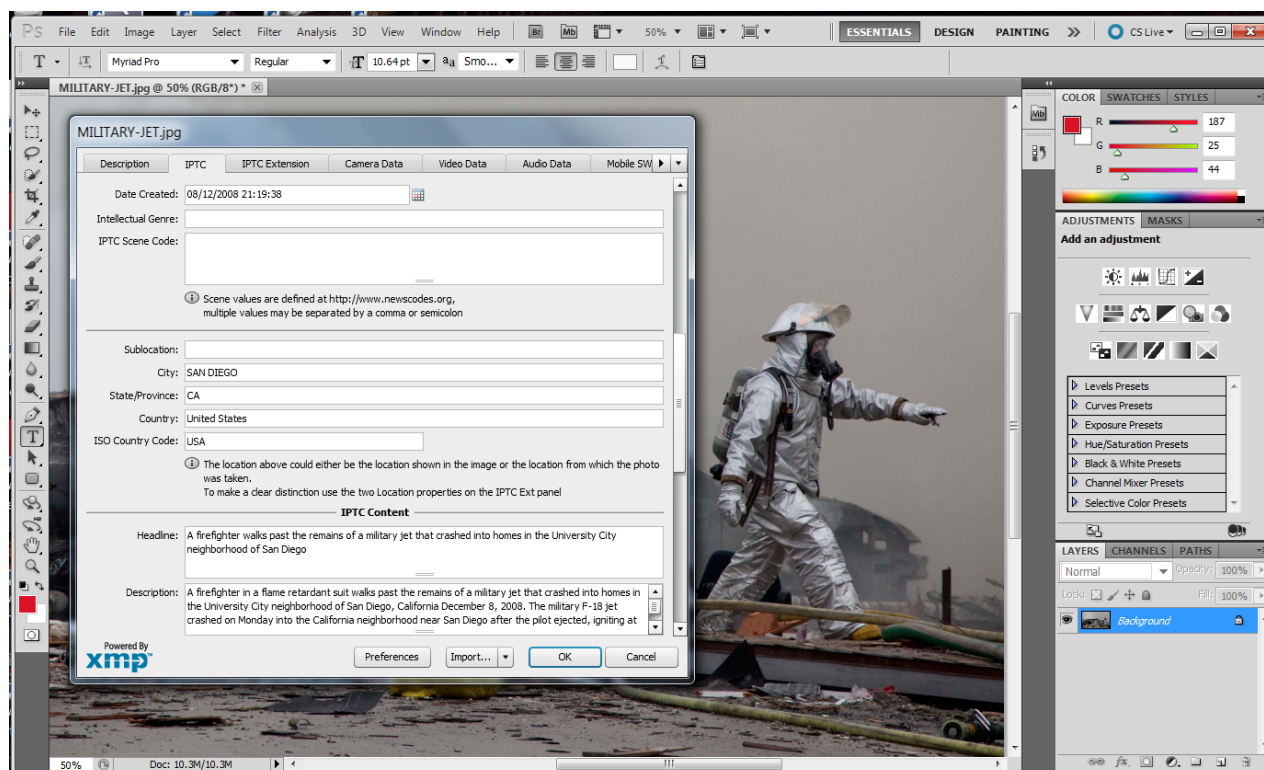


Figure 22: IPTC fields in Photoshop File Info panel (photo and metadata ©Thomson Reuters)

DataSet	IIM Name	Value	Ref
2:05	Object Name	USA-CRASH/MILITARY	22.9.1
2:10	Urgency	4	22.9.2
2:15	Category	A	22.9.3
2:20	Supplemental Category	DEF DIS	22.9.3
2:25	Keywords	Us	22.9.3
2:25	Keywords	Military	
2:25	Keywords	Aviation	
2:25	Keywords	Crash	
2:25	Keywords	Fire	22.9.5
2:40	Special Instruction	NO ARCHIVE OR WIRELESS USE	
2:55	Date Created	20081208	22.9.6



DataSet	IIM Name	Value	Ref
2:60	Time Created	133000-0800	22.9.6
2:80	By-Line	John Smith	22.9.7
2:85	By-line Title	Staff Photographer	22.9.7
2:90	City	San Diego	22.9.8
2:92	Sub-location	University City	22.9.8
2:95	Province/State	CA	22.9.8
2:100	Country/Primary Location Code	USA	22.9.8
2:101	Country/Primary Location Name	United States	22.9.8
2:103	Original Transmission Reference	SAD02	22.9.9
2:105	Headline	A firefighter walks past the remains of a military jet that crashed into homes in the University City neighborhood of San Diego	22.9.10
2:110	Credit	Acme/John Smith	22.9.7
2:115	Source	Acme News LLC	22.9.7
2:116	Copyright Notice	© Copyright Acme News LLC. Click For Restrictions - http://www.acmenews.com/legal	22.9.11
2:120	Caption/Abstract	A firefighter in a flame retardant suit walks past the remains of a military jet that crashed into homes in the University City neighborhood of San Diego, California December 8, 2008. The military F-18 jet crashed on Monday into the California neighborhood near San Diego after the pilot ejected, igniting at least one home, officials said.	22.9.12
2:122	Writer/Editor	CK	22.9.13

22.9.1 Object Name / Title

The <title> element is a child of <itemMeta> and is intended to be a human-readable identifier. The <itemMeta> block also contains some mandatory G2 elements, as shown:

```
<itemMeta>
  <itemClass qcode="ni nat: picture" />
  <provider qcode="nprov: reuters" />
  <versionCreated>2010- 10- 19T02: 20: 00Z</versionCreated>
  <pubStatus qcode="stat: usable" />
  <fileName>USA- CRASH- MI LI TARY_001_HI . j pg</fileName>
  <title>USA- CRASH/MI LI TARY</title>
  <edNote />
</itemMeta>
```

Some providers may additionally map this DataSet to the NewsML-G2 <slugline> property in <contentMeta>:

```
<slugline>USA- CRASH/MI LI TARY</slugline>
```

22.9.2 Urgency

Although the DataSet Supplemental Category were deprecated in the latest version of the IIM, and Urgency and Category in XMP, they are still being used in some services, and therefore if present they may be mapped to the corresponding NewsML-G2 property, which is a child of <contentMeta>:

```
<urgency>4</urgency>
```



22.9.3 Category, Supplementary Category

These IIM DataSets, if present, may be mapped to <subject>. If the values are part of a scheme maintained by the provider, they can be globally and uniquely identified using a QCode:

```
<subject type="cpnat: abstract" qcode="mycat: A" />
<subject type="cpnat: abstract" qcode="mysuppcat: DEF" />
<subject type="cpnat: abstract" qcode="mysuppcat: DIS" />
```

Some providers use Supplemental Category DataSets, which are repeatable in IIM, to hold a type of keyword. Implementers are therefore advised to use the suggested keyword “rules” discussed below.

Note, however, that if these Supplemental Categories are part of a controlled vocabulary, this cannot be expressed using <keyword>, which only allows a string value. Any CV values should be identified using QCodes.

22.9.4 Keywords

The semantics of keyword are somewhat open: some providers use keywords to denote “key” words that can be used by text-based search engines; some use “keyword” to categorise the content using mnemonics, amongst other examples. This makes it difficult to generalise on the correct mapping of the Keyword DataSet to G2.

Rather than assume that the contents of the IIM Keywords DataSet be mapped to NewsML-G2 <keyword>, the IPTC suggests the following rules when configuring a mapping of Keywords metadata:

- ❖ Assess if any existing NewsML-G2 properties align to this use of keyword. Typical examples are
- ❖ Genres (“Feature”, “Obituary”, “Portrait”, etc.)
- ❖ Media types (“Photo”, “Video”, “Podcast” etc.)
- ❖ Products/services by which the content is distributed
- ❖ If the keyword expresses the subject of the content it could go into the <subject> property with the keyword string placed in a <name> child element of the subject with a language tag if required.
- ❖ If migrated to <subject>, providers should also consider:
- ❖ Adding @type if the nature of the concept expressed by the keyword can be determined
- ❖ Using a QCode if there is a corresponding concept in a controlled vocabulary
- ❖ If none of the above conditions are met, then implementers should default to using <keyword> with a @role if possible to define the semantic of the keywords.

The contents of the Keywords field in the example shown have blurred application: they could properly be regarded as subjects, but the provider seems to intend that they be used as natural-language “key” words that could be used by a text-based search engine to index the content. Therefore, they will be mapped to <keyword>.

```
<keyword role="krole: index">us</keyword>
<keyword role="krole: index">military</keyword>
<keyword role="krole: index">aviation</keyword>
<keyword role="krole: index">crash</keyword>
<keyword role="krole: index">fire</keyword>
```

In IIM, the Keywords DataSet is repeatable, with each holding one keyword; therefore each keyword is mapped to separate <keyword> properties, even though they may appear as a comma-separated list in software application dialogs.

22.9.5 Special Instruction

The contents of this field could go into <edNote>, a child of <itemMeta>, which is placed after the <title> element (if present), if the nature of the instruction is a generic message to the receiver or its nature is unknown:

```
<itemMeta>
...
<edNote>NO ARCHIVAL OR WIRELESS USE</edNote>
</itemMeta>
```

If appropriate and advised by the provider, an alternative mapping for the contents of this field MAY be <usageTerms>, parts of the <rightsInfo> block:

```
<rightsInfo>
  <copyrightHolder qcode="prov: AcmeNews">
    <name>Acme News LLC</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright Acme News LLC 2008. For conditions of use see
    http://www.acmenews.com/legal
  </copyrightNotice>
  <usageTerms>NO ARCHIVAL OR WIRELESS USE</usageTerms>
</rightsInfo>
```

22.9.6 Date Created, Time Created

These map to <contentCreated>, a child of <contentMeta>, since it refers to the content itself, for example:

```
<contentMeta>
  <contentCreated>2008- 12- 08</contentCreated>
  ....
</contentMeta>
```

When there is a Time Created value present in the IIM Record, this should be merged with Date Created, as the NewsML-G2 property accepts a Truncated DateTime value (i.e. the value may be truncated if parts of the Date-Time are not available).¹⁵

```
<contentMeta>
  <contentCreated>2008- 12- 08T13: 30: 00- 08: 00</contentCreated>
  ....
</contentMeta>
```

22.9.7 By-line, Credit, Source

These three IIM DataSets are complementary, but have distinct application:

- ❖ By-line is intended to identify the creator of the content
- ❖ Credit identifies the provider of the content
- ❖ Source holds the original owner of the rights to the intellectual property of the content.

The recommended mapping for By-line (IIM 2:80) is to the <creator> child element of <contentMeta>, rather than <by>. This is because <creator> is an administrative property that is intended to be machine-readable; the IPTC recommends that controlled vocabularies should be used if possible. The NewsML-G2 <by> property is a human-readable natural language property that is intended for display, but does not unambiguously identify the creator.

The example below shows this identification metadata in its administrative context. Expressed in this way using QCodes, the metadata can be used for administration and search. Using a CV, the photographer can be uniquely and unambiguously identified. The optional <name> is shown, and the <creator> property also allows the use of the child element <related> which in this case is used to express the photographer's job title, again using a QCode, from IIM DataSet 2:85 (By-line Title)

```
<creator role="crol: photog" qcode="pers: JS001">
  <name>John Smi th</name>
  <related rel="personrel: jobtitle" qcode="stafftitles: photo">
    <name>Staff Photographer</name>
  </related>
</creator>
```

Credit should be mapped to the <creditline> child property of <contentMeta>. There is a <provider> property of <itemMeta>, but the Credit does necessarily reflect the provider. Many picture providers use IIM Credit to display the name of the person, organisation, or both, who should be credited when the picture is used. In this context, <creditline> is appropriate because it is a natural-language label that is intended to be displayed.

```
<creditline>Acme/John Smi th</creditline>
```

¹⁵ If present, the time part must be used in full, with time zone, and ONLY in the presence of the full date.

Source in the IIM specification refers to the initial holder of the copyright, and should therefore be mapped to the <copyrightHolder> child of <rightsInfo>, using a QCode to unambiguously identify the party holding the copyright to the content. However, implementers should be aware of potential issues in the mapping of Source.

In some distribution systems, the original owner of the copyright (as distinct from the current owner) is important, and some providers use the Source field for this information. This is the intention of the IIM Specification which uses Source for the Original Owner, and the Copyright Notice (DataSet 2:116) to hold the copyright statement which includes the Current Owner. Other providers may not use this convention: either the original copyright owner coincides with the current owner, or no distinction is made.

NewsML-G2 provides a single copyrightHolder property, which is intended to contain the Current Owner of the copyright, and this value should take precedence over the Original Owner reflected by the Source property, if different.

```
<rightsInfo>
  <copyrightHolder qcode="prov:AcmeNews">
    <name>Acme News LLC</name>
  </copyrightHolder>
  ...
```

22.9.8 City, Province/State and Country

As discussed in [Quick Start: Pictures and Graphics](#) geographical metadata in images may have different contexts:

- ❖ The location from which the content originates, i.e. where the camera was located. NewsML-G2 has a <located> property to express this.
- ❖ The location shown in the picture – in NewsML-G2 this is a <subject> property of the picture.

Although for the majority of pictures, these are effectively the same spot, one can envisage situations where these two semantically distinct locations are not in the same place: a picture of Mount Fuji taken from downtown Tokyo is one example which is often quoted.

When mapping from IIM or IIM-based Photoshop fields, we assume that the intention of the provider is to express the location shown in the image, as this is the more customary use of these fields. (Be aware that specific providers may have their own convention, and receivers are advised to check).

The <subject> property can have child elements of <name>, and three members of the Concept Definition Group (definition, related, note) and from the Concept Relationships Group (broader, narrower, related, sameAs).

The location shown in the example image could be expressed using a series of separate <subject> properties linked using <broader> to build a subject location hierarchy with “United States” at the top level, and @type is used to hint at the kind of concept identified in the property, for example that it is a city.

An alternative would be to use a single <subject> property for the location, linked to an <assert> block:

```
<contentMeta>
  ...
  <subject type="cpnat:poi" literal="int001" />
  ...
</contentMeta>
<assert literal="int001">
  <type qcode="cpnatexp:geoAreaSublocation" />
  <name>University City</name>
  <broader qcode="mycity:int002" type="cpnatexp:geoAreaCity">
    <name>San Diego</name>
  </broader>
  <broader qcode="mystate:int003" type="cpnatexp:geoAreaProvState">
    <name>California</name>
  </broader>
  <broader qcode="iso3166-1a3:USA" type="cpnatexp:geoAreaCountry">
    <name>United States</name>
  </broader>
</assert>
```



Using <assert> in this way can be an advantage if a concept is used more than once in the NewsML-G2 Item. For example if both the Location Shown and the Location Created are the same place, all of the required concept details can be grouped in one <assert> and shared by both properties.

22.9.9 Original Transmission Reference

This is defined in IIM as “a code representing the location of original transmission”, but in common usage this DataSet has a broader use as an identifier for the purpose of improved workflow handling (IPTC Core: Job ID). These uses include:

- ❖ As an identifier for the picture (perhaps on some content management system)
- ❖ As an identifier for a series of pictures, of which this one is part, e.g. a group of pictures of the same event.

The first use should be mapped to the NewsML-G2 property <altId>, a child of <contentMeta> which is available at PCL. <altId> has two attributes: @type indicates the context of the identifier using a QCode, and @environment indicates the business environment in which the identifier can be used. This is expressed using one or more QCodes (QCode List).

```
<altId type="idtype:systemRef" environment="acmesys:mdn acmesys:iim">SAD02</altId>
```

If the DataSet represents a Job ID, the recommended mapping is to the NewsML-G2 <memberOf> property, a child of <itemMeta> (PCL only):

```
<memberOf type="myref:jobref">
  <name>SAD02</name>
</memberOf>
```

22.9.10 Headline

Maps to the <headline> child of <contentMeta>, a block type element:

```
<headline xml:lang="en-US">A firefighter walks past the remains of a military
jet that crashed into homes in the University City neighborhood of San Diego
</headline>
```

Note the use of @xml:lang to declare the language and variant “en-US”.

22.9.11 Copyright Notice

These fields correspond to the <copyrightNotice> element, a child of the <rightsInfo> block.

```
<copyrightNotice>(c) Copyright Acme News LLC 2008. For conditions of use see
http://www.acmenews.com/legal
</copyrightNotice>
```

22.9.12 Caption/Abstract

The contents of this field are placed in the <description> element, part of <contentMeta> with a @role attribute to denote that the description is a picture caption:

```
<description xml:lang="en-US" role="drol:caption">A firefighter in a flameproof
suit walks past the remains of a military jet that crashed into homes in the
University City neighborhood of San Diego, California December 8, 2008. The
military F-18 jet crashed on Monday into the California neighborhood near San
Diego after the pilot ejected, igniting at least one home, officials said.
</description>
```

22.9.13 Writer/Editor

The caption writer is often a different person to the photographer, so aligns with the NewsML-G2 <contributor> property, a child of <contentMeta>. If possible, use a QCode value to unambiguously identify the contributing person and a @role to describe their role in the workflow. This property may be extended (at PCL) to include contact details and other information such as job title.

```
<contributor role="crol:capwriter" qcode="pers:CK" />
```




LISTING 26 Embedded photo metadata fields mapped to NewsML-G2

The listing below combines the examples above into a complete listing. The following options have been used:

- ❖ <usageTerms> used for Special Instructions, instead of <edNote>
- ❖ <assert> used in conjunction with <subject> to express the location shown in the image.
- ❖ <altId> used for the Original Transmission Reference (instead of <memberOf>)

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:acmenews.com,2008:WORLD-NEWS:USA20090209098658"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef href="http://www.acmenews.com/customer/cv/catalog4customers-1.xml" />
  <rightsInfo>
    <copyrightHolder qcode="prov:AcmeNews">
      <name>Acme News LLC</name>
    </copyrightHolder>
    <copyrightNotice>(c) Copyright Acme News LLC 2008. For conditions of use see
      http://www.acmenews.com/legal
    </copyrightNotice>
    <usageTerms>NO ARCHIVAL OR WIRELESS USE</usageTerms>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ni nat: picture" />
    <provider qcode="nprov: reuters" />
    <versionCreated>2010-10-19T02:20:00Z</versionCreated>
    <pubStatus qcode="stat: usable" />
    <fileName>USA-CRASH-MILITARY_001_HI.jpg</fileName>
    <title>USA-CRASH/MILITARY</title>
    <edNote />
  </itemMeta>
  <contentMeta>
    <urgency>4</urgency>
    <contentCreated>2008-12-08T13:30:00-08:00</contentCreated>
    <creator role="crol: photog" qcode="pers: JS001">
      <name>John Smith</name>
      <related rel="personrel: jobtitle" qcode="stafftitles: photo">
        <name>Staff Photographer</name>
      </related>
    </creator>
    <altId type="idtype: systemRef" environment="acmesystem: mdn: iim">SAD02</altId>
    <contributor role="crol: capwriter" qcode="pers: CK" />
    <subject type="cpnat: abstract" qcode="mycat: A" />
    <subject type="cpnat: abstract" qcode="mysuppcat: DEF" />
    <subject type="cpnat: abstract" qcode="mysuppcat: DIS" />
    <subject type="cpnat: poi" literal="int001" />
    <headline xml:lang="en-US">A firefighter walks past the remains of a military
      jet that crashed into homes in the University City neighborhood of San Diego
    </headline>
    <description xml:lang="en-US" role="drol: caption">A firefighter in a flameproof
      suit walks past the remains of a military jet that crashed into homes in the
      University City neighborhood of San Diego, California December 8, 2008. The
      military F-18 jet crashed on Monday into the California neighborhood near San
      Diego after the pilot ejected, igniting at least one home, officials said.
    </description>
    <creditline>Acme/John Smith</creditline>
    <keyword role="krol: index">us</keyword>
    <keyword role="krol: index">military</keyword>
    <keyword role="krol: index">aviation</keyword>
    <keyword role="krol: index">crash</keyword>
    <keyword role="krol: index">fire</keyword>
    <slugline>USA-CRASH/MILITARY</slugline>
  </contentMeta>
  <assert literal="int001">
    <type qcode="cpnatexp: geoAreaSublocation" />
    <name>University City</name>
    <broadener qcode="mycity: int002" type="cpnatexp: geoAreaCity">
      <name>San Diego</name>
    </broadener>
    <broadener qcode="mystate: int003" type="cpnatexp: geoAreaProvState">
      <name>California</name>
    </broadener>
  </assert>
</newsItem>
```



```
<broadcaster qcode="iso3166-1a3:USA" type="cpnatexp:geoAreaCountry">
  <name>United States</name>
</broadcaster>
</assert>
<contentSet>
  <remoteContent
    href="http://www.acmenews.com/content/pictures/hi res/20090209/USA-CRASH-
    MILITARY_001_HI.jpg"
    rendition="rnd:highRes" size="3764418" contentType="image/jpeg" width="2500"
    height="1445" colourSpace="colsp:USWOPv2" />
  </contentSet>
</newsItem>
```

22.10 Reconciling NewsML-G2 with embedded metadata

Where equivalent properties exist in both NewsML-G2 metadata and embedded (XMP or EXIF) metadata, the IPTC recommends that embedded administrative metadata, such as <located>, MAY take precedence over NewsML-G2 metadata, subject to guidance from the provider, and always with caution.

For example, a picture may have GPS metadata embedded by the camera, which may be different from the <located> property entered by a journalist.

The difference may be one of precision: the GPS co-ordinates may be precise, but hardly useful to the ultimate consumer. Even if a human-readable value is derived from the GPS data, would this be better than the information, if accurate, provided by the journalist? If the difference is due to inaccuracy, the receiver would have no way of knowing whether the journalist has made a mistake, or whether the camera is incorrectly set.



This page intentionally blank



23 Rights Metadata

Rights-related metadata has become increasingly important, and feedback to the IPTC from the news industry pointed to the need for a more granular design of applying `<rightsInfo>` to the item and/or content.

The IPTC has taken over responsibility for the management of the ACAP (Automated Content Access Protocol), the publisher-led project to develop tools for the communication of copyright permissions information on web sites. As a next step a news content specific Profile for Open Digital Rights Language (ODRL) version 2.0 has been defined, branded RightsML, which identifies those features of ODRL 2.0 that can most usefully be employed in news syndication and defines additional vocabulary for expressing the usage permissions and constraints that are particular to the news industry. RightsML enables a wide range of usage policies to be expressed in NewsML-G2. See the RightsML web site www.rightsml.org for further details.

While the NewsML-G2 `<rightsInfo>` wrapper and its properties are machine-readable, the assertion of rights within the NewsML-G2 properties is expressed in natural-language terms. However, the Extension Point of `<rightsInfo>` enables properties from another namespace to be added, and this would include ODRL and the RightsML Profile. A description of this implementation is outside the scope of these Guidelines.

Machine-reading of `<rightsInfo>` by a processor is targeted at being able to determine the following:

- ❖ To which element(s) inside an item does a specific `<rightsInfo>` wrapper apply? (rights holder's view)
- ❖ Which is/are the `<rightsInfo>` wrapper(s) that apply to a specific element? (content user's view)

These two processing cases are covered in [23.6](#) below.

23.1 Rights Info (Core Conformance Level)

The `<rightsInfo>` wrapper has child elements of

- ❖ `accountable` (0..1) The person responsible for legal issues regarding the content (in some jurisdictions)
- ❖ `copyrightHolder` (0..1) The person or organisation claiming intellectual property ownership of the content
- ❖ `copyrightNotice` (0..unbounded) The formal statement claiming the Intellectual Property Ownership (IPO)
- ❖ `usageTerms` (0..unbounded) A natural-language statement of how the content may be used.

For example:

```
<rightsInfo>
  <copyrightHolder qcode="nprov:reuters" />
  <name>Reuters</name>
  <copyrightNotice>(c) Copyright Reuters 2010. Click For
    Restrictions - http://about.reuters.com/fulllegal.asp
  </copyrightNotice>
  <usageTerms xml:lang="en">NO ARCHIVAL OR WIRELESS USE</usageTerms>
</rightsInfo>
```

23.2 Link to an external Rights Information resource

The G2 Rights Information wrapper accommodates most cases where a news provider needs to communicate the rights associated with a G2 Item. Although the structure is machine-readable, the information it contains is intended to be read and interpreted by people.

Machine-readable rights metadata encoded as XML, for example using RightsML, may be added directly to a G2 document via the Extension Point to `<rightsInfo>`. As an alternative a provider may need to refer to an external resource containing rights information and instructions. This can be accomplished using the `<link>` element of `<rightsInfo>`:

```
<rightsInfo>
  <copyrightHolder uri="http://www.acmenews.com" />
  <copyrightNotice>(c) 2012 - Acme Newswire</copyrightNotice>
```

```
<link
  href="http://acmenews.com/rights/rightspolicy123.xml"
  contentType="text/xml"
/>
</rightsInfo>
```

23.3 Business Reasons for Extended Rights Information at PCL

- ❖ Business users potentially need to assert different rights to different parts of a News Item. For example, a photo library may not own the copyright to a picture, but does want to assert intellectual property ownership of the keywords (i.e. the metadata). This leads to the need to distinguish between rights to the *content*, and rights to *metadata*, or *parts of the metadata*. End users need a simple mechanism for addressing a whole group of XML elements as being governed by a `<rightsInfo>` expression.
- ❖ Implementers need a clear set of rules for expressing rights to XML elements identified by an `@id`. The `<partMeta>` wrapper is an existing NewsML-G2 construct for this purpose and in order to have a consistent processing model, the `<rightsInfo>` wrapper expresses rights via `<partMeta>` to the target element, rather than directly to the target element.

23.4 The extended `<rightsInfo>` model

The optional `@idrefs` attribute is defined as:

“Reference(s) to the part(s) of an Item to which the `<rightsInfo>` element applies. When referencing part(s) of the content of an Item, `@idrefs` *must* include the `@partid` value of a `partMeta` element which in turn references the part of the content.”

Two attributes enable implementers to define the scope of a rights expression and assert the aspect of the intellectual property rights that are being claimed:

- ❖ `scope`; (0..1), `QCodeListType`; allows a provider to indicate whether the rights expression is about the content, the metadata, or both. If the attribute is not present then `<rightsInfo>` applies to all parts of the Item. There is a mandatory NewsCodes scheme for the values, see below. The CV contains two values: “content” and “metadata”.
- ❖ `aspect`; (0..1), `QCodeListType`; allows the provider to assert rights to a specific intellectual effort. In terms of metadata, this can express rights to the actual metadata values used, and/or it can express rights to the work in applying metadata values. `@aspect` also enables providers to assert rights to the structure of content, for example the structure of a package of news. If the attribute does not exist then `<rightsInfo>` applies to all aspects. There is a mandatory NewsCodes scheme for the values, see below. The CV contains three values: “values”, “selection” and “structure”.

23.4.1 Rights Info Scope NewsCodes

Scheme URI: <http://cv.iptc.org/newscodes/riscope/>

Recommended scheme alias: riscope

Code	Definition
Content	Encompasses all child elements of a) <code><contentSet></code> in a News Item, b) <code><concept></code> in a Concept Item, c) <code><conceptSet></code> in a Knowledge Item and d) <code><groupSet></code> (except all children of <code><itemRef></code>) in a Package Item. Any parts of content which are described by <code><partMeta></code> elements are also included.
Metadata	Encompasses all child elements of an item, except a) those that are in the scope of the 'content' scope indicator and b) all children of <code><itemRef></code> elements in a Package Item.

23.4.2 Rights Aspect NewsCodes

Scheme URI: <http://cv.iptc.org/newscodes/raspect/>

Recommended scheme alias: raspect

Code	Definition
Values	The <code><rightsInfo></code> element makes an assertion about metadata values including the details of concepts. This aspect applies only to the Rights Info Scope of

Code	Definition
	"metadata".
Selection	The <rightsInfo> element makes an assertion about selecting and applying metadata values or selecting and applying item references in a Package Item. This aspect applies a) to the Rights Info Scope of "metadata" and b) to the Rights Info Scope of "content" for Package Items.
Structure	The <rightsInfo> element makes an assertion about the design of the structure of the content, e.g. the structure of a Package Item. This aspect applies only to the Rights Info Scope of "content".

23.5 Use cases for the extended <rightsInfo> model

23.5.1 Separating rights for "content" and "metadata"

A photo library sends a customer a picture. The Intellectual Property (IP) in the picture itself (the **content**) is owned by a third party. The photo library wants to assert the rights to the **metadata** that accompanies the picture. This is expressed by the following <rightsInfo> blocks:

```
<newsItem> ...
  <!-- Content: Third Party -->
  <rightsInfo scope="rscope:content" >
    <copyrightHolder>
      <name>Example Pictures</name>
    </copyrightHolder>
    <copyrightNotice>(c) Copyright Example Pictures Ltd 2008. Click For
      Restrictions - http://about.example.com/legal.asp
    </copyrightNotice>
    <usageTerms xml:lang="en">MUST COURTESY PARAMOUNT PICTURES
      FOR USE OF "THE CURIOUS CASE OF BENJAMIN BUTTON" WITH NO ARCHIVAL USE
    </usageTerms>
  </rightsInfo>
  <!-- Metadata: Photo Library -->
  <rightsInfo scope="rscope:metadata" >
    <copyrightHolder>
      <name>The Picture Library</name>
    </copyrightHolder>
    <copyrightNotice>(c) Copyright The Picture Library Ltd 2008.
    </copyrightNotice>
  </rightsInfo>
```

23.5.2 Separating rights for part of the metadata

As in **23.5.1** the photo library sends a third-party picture (the **content**), but wishes to express rights to only **part of the metadata**. The metadata properties are identified by @idrefs, and the scope has been omitted, since the target element is a metadata property (see what a "metadata property" is in **23.4.1**).
[copyrightHolder!]

```
<newsItem>
  ...
  <!-- All: Third Party -->
  <rightsInfo>
    <copyrightHolder>
      <name>Example Pictures</name>
    </copyrightHolder>
    <copyrightNotice>(c) Copyright Example Pictures Ltd 2008. Click For
      Restrictions - http://about.example.com/legal.asp
    </copyrightNotice>
    <usageTerms xml:lang="en">MUST COURTESY PARAMOUNT PICTURES
      FOR USE OF "THE CURIOUS CASE OF BENJAMIN BUTTON" WITH NO ARCHIVAL USE
    </usageTerms>
  </rightsInfo>
  <!-- Part of the Metadata: Picture Library -->
  <rightsInfo idrefs="id001 id002 id003 id004 id005">
    <copyrightHolder>
      <name>The Picture Library</name>
    </copyrightHolder>
    <copyrightNotice>(c) Copyright The Picture Library Ltd 2008.
    </copyrightNotice>
  </rightsInfo>
  ...
</contentMeta>
```

```
...
<keyword id="id001" role="krole:index">us</keyword>
<keyword id="id002" role="krole:index">military</keyword>
<keyword id="id003" role="krole:index">aviation</keyword>
<keyword id="id004" role="krole:index">crash</keyword>
<keyword id="id005" role="krole:index">fire</keyword>
...
</contentMeta>
...
```

23.5.3 Separating rights for aspects of part of the metadata

As in [23.5.2](#) the photo library sends a third-party picture (the *content*), expressing rights to *part of the metadata*. Since the IP of the scheme being used belongs to another party, the picture library expresses its ownership of rights associated with *applying* the metadata to the picture, and also acknowledges the Scheme Authority's rights to the code values themselves.

```
<newsItem>
...
<!-- All: Third Party -->
<rightsInfo>
  <copyrightHolder>
    <name>Example Pictures</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright Example Pictures Ltd 2008. Click For
    Restrictions - http://about.example.com/legal.asp
  </copyrightNotice>
  <usageTerms xml:lang="en">MUST COURTESY PARAMOUNT PICTURES
    FOR USE OF "THE CURIOUS CASE OF BENJAMIN BUTTON" WITH NO ARCHIVAL USE
  </usageTerms>
</rightsInfo>
<!-- Part of the Metadata Values: Another Party -->
<rightsInfo idrefs="id001 id002" aspect="aspect:values">
  <copyrightHolder>
    <name>International Press Telecommunications Council</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright International Press Telecommunications
    Council 2008.
  </copyrightNotice>
</rightsInfo>
<!-- Part of the Metadata Selection: Picture Library -->
<rightsInfo idrefs="id001 id002" aspect="aspect:selection">
  <copyrightHolder>
    <name>The Picture Library</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright The Picture Library Ltd 2008.
  </copyrightNotice>
</rightsInfo>
...
<contentMeta>
...
  <subject id="id001" type="cpnat:abstract" qcode="medtop:20000106" />
  <subject id="id002" type="cpnat:abstract" qcode="medtop:02000000" />
...
</contentMeta>
...
```

23.5.4 Expressing rights to different parts of content

A news agency sends a text article (the *content*) that includes a substantial *extract of content* owned by another party. The overall rights to both the content and metadata of the article are expressed using a `<rightsInfo>` wrapper with no attributes. The rights owned by the third party to the extract are "carved out" by using a separate `<rightsInfo>` wrapper that references the extract within the article via `<partMeta>`:

```
<newsItem>
...
<!-- All: News Agency -->
<rightsInfo>
  <copyrightHolder>
    <name>Example News</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright Example News Ltd 2010. Click For
    Restrictions - http://about.example.com/legal.asp
  </copyrightNotice>
  <usageTerms xml:lang="en">

```

```

    REGIONAL PRESS ONLY. NO BROADCAST OR WEB USE
  </usageTerms>
</rightsInfo>
<!-- Part of the Content: Another Party -->
<rightsInfo idrefs="id001">
  <copyrightHolder qcode="nprov: reuters">
    <name>Thomson Reuters</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright Thomson Reuters 2010.
  </copyrightNotice>
</rightsInfo>
...
</contentMeta>
<partMeta partid="id001" contentRefs="p001"/>
</inlineXML>
...
<p>...still the aftershocks of the banking crisis continue to rumble,
  With many analysts noting big changes in the role of sovereign wealth
  funds.
  ... </p>
<p>According to a recent special report by Thomson Reuters: <br />
<span id="p001">"Dubai's crisis prompted a shift of power to the rulers in Abu
  Dhabi, the wealthiest of the seven states that make up the United Arab
  Emirates. Now a chastened Dubai is recovering some of its confidence as it
  seeks to convince international investors it can deliver now where last
  year it failed."</span></p>
...
</inlineXML>

```

23.5.5 Expressing rights to aspects of content

A news provider operates a distribution platform where third-party companies can sign up to provide their own news and information packages to the subscriber base. The platform provider has created standard package templates using NewsML-G2, which the other companies populate with their chosen topics. The provider asserts its right to the *structure* of the news packages, whilst the rights to the *selection* of the news belong to the third party, as shown:

```

<packageItem>
<!-- A Package Item is a selection of references to G2 Items organised into -->
<!-- a structure of Groups, each with a @role, within a Group Set. -->
...
<!-- Structure: News Provider -->
<rightsInfo scope="riscope: content" aspect="raspect: structure">
  <copyrightHolder>
    <name>The News Provider</name>
  </copyrightHolder>
  <copyrightNotice>
    (c) Copyright asserted to the structure of the package <groupSet>
  </copyrightNotice>
  <usageTerms xml:lang="en">
    Usage terms in natural language
  </usageTerms>
</rightsInfo>
<!-- Selection: Third Party -->
<rightsInfo scope="riscope: content" aspect="raspect: selection">
  <copyrightHolder>
    <name>The News Selector</name>
  </copyrightHolder>
  <copyrightNotice>
    (c) Copyright asserted to the selection of the Item References.
  </copyrightNotice>
  <usageTerms xml:lang="en">
    Usage terms in natural language
  </usageTerms>
</rightsInfo>
<!-- Metadata: News Provider -->
<rightsInfo scope="riscope: metadata" >
  <copyrightHolder>
    <name>The News Provider</name>
  </copyrightHolder>
  <copyrightNotice>
    (c) Copyright to the package metadata asserted here.
  </copyrightNotice>
</rightsInfo>
...
<groupSet root="root">
  <group id="root" role="group: root" mode="pgrmod: bag">

```



```
<groupRef idref="grp1" />
<groupRef idref="grp2" />
</group>
<group id="grp1" role="group: main">
  <itemRef ...>
    ...
  </itemRef>
  <itemRef ...>
    ...
  </itemRef>
</group>
<group id="grp2" role="group: si debar">
  <itemRef ...>
    ...
  </itemRef>
  <itemRef ...>
    ...
  </itemRef>
</group>
</groupSet>
<packageItem>
```

Note that the rights expressed about the package *content* selection and *structure* are NOT inherited by the Items referenced by the package in each <itemRef>.

23.6 Processing models for extended Rights Info

Please see the Rights Information chapter of the NewsML-G2 Specification, which may be downloaded by visiting www.newsml-g2.org and navigating to Specification resources.

24 Expressing company financial information

24.1 Background

Company financial information is a common feature of news distributed by many news providers; company information is typically indexed in news using a “ticker”, for example:

Apple Inc (NYSE:AAPL) today announced the...

However, this easy and widely-used “shorthand” for indexing company information has a number of traps for those creating machine-readable metadata, especially when dealing with global markets and diverse customer requirements.

Tickers are NOT unambiguous identifiers for companies; they identify specific financial instruments belonging to companies, usually tied to specific markets or exchanges. For example, the ticker for Rio Tinto is “RIO” in London, but “RTP” in New York; similarly NYSE:AAPL identifies shares of Apple Inc. traded on the New York Stock Exchange, not the company. Other issues include:

- ❖ A company may have many different financial instruments, all identified by specific tickers
- ❖ A company’ shares may be traded on many different markets; the ticker symbols may or may not be the same across markets. For example IBM has the same Ticker Symbol IBM on the New York, Amsterdam, Frankfurt, London and Mexico Exchanges; as mentioned above, Rio Tinto has the Ticker Symbol RIO in London, but RTP on the New York, Frankfurt and Mexico exchanges.
- ❖ There are many different schemes for labelling financial markets, for example the London Stock Exchange is variously identified as “LON”, “LSE”, “LN” to give only three examples.

The G2 <hasInstrument> structure addresses these issues by enabling providers to use a globally unique identifier for companies, and linking each of the company tickers to this identifier, while giving extensive and unambiguous information about each of the tickers. It is available at PCL as a child of the Organisation Details wrapping element <organisationDetails>

24.2 <hasInstrument> properties (as attributes)

Definition	Name	Cardinality	Datatype	Example/Notes
A symbol for the financial instrument	symbol	(1)	String	symbol="RIO"
The source of the financial instrument symbol	symbolsrc	(0..1)	QCodeType	symbolsrc="symsrc:MDNA"
A venue in which this financial instrument is traded	market	(0..1)	QCodeType	market="mic:XLON" The scheme alias “mic” refers to the ISO-supported Market Identification Code scheme
The label used for the market	marketlabel	(0..1)	String	marketlabel="LSE"
The source of the market label	marketlabelsrc	(0..1)	QCodeType	marketlabelsrc="mlsrc:MDNA"
The type(s) of the financial instrument	type	(0..1)	QCodeListType	type="instrtype:share"

24.2.1 Examples

At its simplest, a symbol and market label may be sufficient

```
<hasInstrument
  symbol = "AAPL"
  marketLabel = "NYSE"
/>
```

If it is important that the authority for the symbol is identified, the optional @symbolsrc may be used:

```
<hasInstrument
  symbol = "AAPL"
  symbolsrc = "symsrc: MDNA"
  marketLabel = "NYSE"
/>
```

Other share identifiers may be used. In this example the ISIN (International Stock Identification Number) is used, and the authority for symbol identified as the ISO. Note that the CV only identifies the authority, not the authority's scheme; which may not be available to the end user. This example does not give a market label, but unambiguously identifies the market on which the instrument is traded using @market and a value from the Market Identification Codes (an ISO scheme) using the scheme alias "mic"

```
<hasInstrument
  symbol = "12345678"
  symbolsrc = "symsrc: ISO"
  marketLabel = "XLON"
/>
```

The @marketlabel enables the provider to denote the market on which the financial instrument is traded from their own or another's scheme, and to identify the scheme using the @marketlabelsrc. For example the following identifies the Toronto Stock Exchange from Bloomberg's CV of markets:

```
<hasInstrument
  symbol = "12345678"
  symbolsrc = "symsrc: ISO"
  marketLabel = "####CN"
  marketLabelsrc = "mlsrc: bbg"
/>
```

24.3 Adding <hasInstrument> to a G2 item

In order to use the <hasInstrument> property inline within a NewsML-G2 News Item, there first needs to be a <subject> property in <contentMeta> (or <partMeta>) that references the company for which financial and other details will be expressed:

```
<contentMeta>
  <contentCreated>2011- 12- 06T07: 55: 00+00: 00</contentCreated>
  <subject type="cpnat: organization" literal="Acme Widget">
    <name>Acme Widget Sales Inc</name>
  </subject>
  <headline>
    Acme Widgets announces annual results
  </headline>
  <description role="ANDescRole: abstract">
    Acme Widgets (XTSE: AWG), a leader in the
    production of quality metal widgets, announced a six per cent increase in
    profits...
  </description>
  <language tag="en-US" />
</contentMeta>
```

In this example, the company is referenced using a @literal value, but either a @qcode or @uri could be used instead.

However, a <subject> element is only a container for referencing a concept, and does not hold extensive details about the concept itself. In order to express further information about the company within the NewsML-G2 document, we use an <assert> wrapper with a reference to the <subject>.

Into this wrapper we can place a further wrapping element <organisationDetails> which can contain the <hasInstrument> information:

```
<assert literal="Acme Widget">
  <organisationDetails>
```



```

    <contactInfo>
      <phone>+1 (416) 922 5834</phone>
      <email>info@acmewidgets.com</email>
    </contactInfo>
    <hasInstrument
      symbol="AWG"
      symbolsrc="symsrc:MDNA"
      market="mic:XTSE"
      marketlabel="TSX"
      marketlabelsrc="mlsrc:MDNA"
      type="instrtype:share" />
    </organisationDetails>
  </assert>

```

For more details on the use of <assert> see [The Assert wrapper](#)

LISTING 27 Company Financial Information

```

<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:acmenews.com:20020408:201003230594296001"
  version="1"
  standard="NewsML-G2"
  standardversion="2.18"
  conformance="power"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
  <catalogRef
    href="http://www.acmenews.com/NewsML/acmenews_NewsML-G2_Catalog_1.xml" />
  <rightsInfo>
    <copyrightHolder uri="http://www.acmenews.com" />
    <copyrightNotice>(c) 2012 - Acme Newswire</copyrightNotice>
    <usageTerms>Not for distribution in the United
      States</usageTerms>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="inat:text" />
    <provider uri="http://www.acmenews.com" />
    <versionCreated>2011-12-06T08:00:00+00:00</versionCreated>
    <firstCreated>2011-12-06T08:00:00+00:00</firstCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <contentCreated>2011-12-06T07:55:00+00:00</contentCreated>
    <subject type="cpnat:organisation" literal="Acme Widget">
      <name>Acme Widget Sales Inc</name>
    </subject>
    <headline>
      Acme Widgets announces annual results
    </headline>
    <description role="ANDescRole:abstract">
      Acme Widgets (XTSE:AWG), a leader in the
      production of quality metal widgets, announced a six per cent increase in
      profits....
    </description>
    <language tag="en-US" />
  </contentMeta>
  <assert literal="Acme Widget">
    <organisationDetails>
      <contactInfo>
        <phone>+1 (416) 922 5834</phone>
        <email>info@acmewidgets.com</email>
      </contactInfo>
      <hasInstrument
        symbol="AWG"
        symbolsrc="symsrc:MDNA"
        market="mic:XTSE"
        marketlabel="TSX"
        marketlabelsrc="mlsrc:MDNA"
        type="instrtype:share" />
      </organisationDetails>
    </assert>
  <contentSet>
    <inlineXML contenttype="application/xhtml+xml">
      <!-- Story -->
    </inlineXML>
  </contentSet>
</newsItem>

```



This page intentionally blank

25 Advanced Metadata Techniques

25.1 Introduction

NewsML-G2 provides powerful tools for adding value and navigation to content, using the <assert> and <inlineRef> properties

In the following sections, we will show how implementers can add features such as text mark-up to highlight people and places in the news and provide navigation to further resources.

25.2 The Assert wrapper

The Assert wrapper is an optional and repeatable child of any root Item (<newsItem>, <conceptItem> etc.) and represents a concept that can be identified using either @uri, @qcode or @literal. Asserts are only available at PCL (@conformance="power")

An Assert creates a one-time instance of supplementary information about a concept that is local to the Item and can be referenced by the Item's other properties using either a @uri, @qcode or @literal identifier.

The other attributes of <assert> are (from the Edit Attributes group):

Name	Datatype
id	XML Schema ID
creator	QCode
modified	DateOptTime

and from the Internationalization group (**i18nAttributes**):

Name	Datatype
xml:lang	XML Schema language
dir	XML Schema string Enumeration <i>ltr</i> , <i>rtl</i>

The content model for <assert> allows it to contain any element from the IPTC's "NAR" namespace, (and from any another namespace). To avoid ambiguity, the rules that **MUST** be followed when adding NewsML-G2 elements to an Assert are:

- ❖ Immediate child elements from <itemMeta>, <contentMeta> or <concept> may be added directly to the Assert without the parent element;
- ❖ All other elements **MUST** be wrapped by their parent element(s), excluding the root element.

25.2.1 Reasons for using Asserts

Sometimes it is impractical, inefficient or undesirable to expect receivers of Items to retrieve rich structured information about concepts from remote resources. In these cases a sub-set of metadata relevant to the Item can be stored directly in the Item using Assert. For example:

- ❖ A provider may decide that it is important to convey supplementary information locally within a News Item, such as how to contact an organisation, that requires a more comprehensive property structure "borrowed" from the Concept wrapper.
- ❖ A provider needs to convey a one-time transient value, such as a company's latest share price, directly within the NewsML-G2 document.
- ❖ A news organisation may wish to add value to stories using supplementary information about concepts that are part of controlled vocabularies, but does not wish to give unrestricted access to the knowledge bases that the information is drawn from.
- ❖ A provider wishes to embed supplementary information about concepts that are NOT part of controlled vocabularies.
- ❖ Performance or connectivity issues may restrict the ability of receivers to retrieve information that is stored remotely.

For example, a Subject property can reference a local Assert, allowing the use of properties that are not directly supported by the <subject> element:

```
<contentMeta>
...
<subject type="cpnat:organisation" literal="Acme Widget">
  <name>Acme Widget Sales Inc</name>
</subject>
...
</contentMeta>
<assert literal="Acme Widget">
  <organisationDetails>
    <contactInfo>
      <phone>+1 (416) 922 5834</phone>
      <email>info@acmewidgets.com</email>
    </contactInfo>
  </organisationDetails>
</assert>
```

This example shows how a concept identified by "Acme Widget" which is a Subject of the Item, can be expanded locally using an Assert.

It illustrates the use of @literal as an identifier whose scope is purely local to the Item. Both @qcode and/or @uri may be used as globally-unique identifiers but note that the concept information contained in an Assert in a Item is ALWAYS local in scope, whatever the type of identifier. Receivers must NOT use the information contained in the <assert> outside the scope of the Item providing the information, for example by extracting it and storing it in a cache of concepts, or in a knowledge base.

This is because the <assert> is intended to be a single-use "snapshot" of information about a concept, if the information needs to be used permanently and periodically updated, the full Concept Item or Knowledge Item, which also convey management metadata, should be used instead.

25.2.1.1 Merging concept information that is used by multiple properties.

If a document, for example, has <subject> and <located> properties that reference a single geographical location (i.e. a picture SHOWS Stonehenge and was taken AT Stonehenge). The <assert> wrapper can provide information about this location that is shared by both properties.

```
<contentMeta>
...
<located type="cpnatexp:geoAreaSublocation" qcode="myalias:int001" />
<!-- Camera location -->
...
<subject type="cpnatexp:geoAreaSublocation" qcode="myalias:int001" />
<!-- Subject of picture -->
...
</contentMeta>
<assert qcode="myalias:int001">
  <type qcode="cpnatexp:geoAreaSublocation" />
  <name>Stonehenge</name>
  <broadener type="cpnatexp:geoAreaCity">
    <name>Amesbury</name>
  </broadener>
  <broadener type="cpnatexp:geoAreaProvState">
    <name>Wiltshire</name>
  </broadener>
  <broadener qcode="iso3166-1a3:GBR" type="cpnatexp:geoAreaCountry">
    <name>United Kingdom</name>
  </broadener>
</assert>
```

25.2.1.1.1 Example: Merged POI Details

In the example below an <address> wrapper has been added as an optional child of <POIDetails> to convey the postal address of the location of the POI. It is recommended that the address wrapped by <contactInfo> should NOT be used to express the location of the POI.

```
<contentMeta>
  <located qcode="artven:int014" />
  <subject type="cpnat:abstract" qcode="medtop:20000030">
    <name>music theatre</name>
  </subject>
  <subject type="cpnat:poi" qcode="artven:int014" />
</contentMeta>
<assert qcode="artven:int014">
```

```
<name>The Metropolitan Opera House</name>
<definition xml:lang="en-US">
  The Metropolitan Opera House is situated in the Lincoln Center for the
  Performing Arts on the Upper West Side of Manhattan, New York City. <br/>
  The Opera House is located at the center of the Lincoln Center Plaza, at
  the western end of the plaza, at Columbus Avenue between 62nd and
  65th Streets. <br/>
</definition>
<POIDetails>
  <address>
    <line>Lincoln Center</line>
    <locality type="geotype:city" literal="int015">
      <name>New York</name>
    </locality>
    <area type="geotype:provstate" literal="int016">
      <name>New York</name>
    </area>
    <country literal="US">
      <name role="nrol:display">United States</name>
    </country>
    <postalCode>10023</postalCode>
  </address>
  <contactInfo>
    <web>http://www.themet.org</web>
  </contactInfo>
</POIDetails>
</assert>
```

25.2.1.2 Using concept details in <assert> in previous versions

Up to NewsML-G2 2.4 and EventsML-G2 1.3, the <concept> wrapper has mandatory elements of <conceptId> and <name>, and therefore omitting these properties would cause errors if validating against these versions of the schema. The workaround to avoid validation errors is to add a “dummy” Concept ID, using a randomly-generated code added to a dummy Scheme URI provided by the IPTC.

Now, immediate child properties of <concept> may be used directly in <assert>, as is already the case for <itemMeta> and <contentMeta>, so the need for this workaround is removed.

25.2.1.2.1 Example: Using the <concept> wrapper with valid ID

In the following example, the <concept> wrapper is used, but the assertion is about a concept from a controlled vocabulary, so the concept ID makes sense. This assertion will contain supplementary information about a concept identified as “artven:int014”: used in a <located> and a <subject> properties of <contentMeta>. (Note that the scope of the information in the <assert> wrapper is LOCAL to the document, and may only be updated when the containing Item is modified.)

The example in more detail:

```
<contentMeta>
  <located qcode="artven:int014" />
  <subject type="cpnat:abstract" qcode="subj:01017001">
    <name>music theatre</name>
  </subject>
  <subject type="cpnat:poi" qcode="artven:int014" />
</contentMeta>
<assert qcode="artven:int014">
  <concept>
    <conceptId qcode="artven:int014" />
    <name>The Metropolitan Opera House</name>
    <definition xml:lang="en-US">
      The Metropolitan Opera House is situated in the Lincoln Center for the
      Performing Arts on the Upper West Side of Manhattan, New York City. <br/>
      The Opera House is located at the center of the Lincoln Center Plaza, at
      the western end of the plaza. <br/>
    </definition>
    <POIDetails>
      <contactInfo>
        <web>http://www.themet.org</web>
      </contactInfo>
      <address>
        <line>Lincoln Center</line>
        <locality type="geotype:city">
          <name>New York</name>
        </locality>
        <area type="geotype:provstate">
          <name>New York</name>
        </area>
      </address>
    </POIDetails>
  </concept>
</assert>
```



```

        </area>
        <country>
          <name role="nrol:display">United States</name>
        </country>
        <postalCode>10023</postalCode>
      </address>
    </contactInfo>
  </POIDetails>
</concept>
</assert>

```

25.2.1.2.2 Example: Using a “dummy” ID

If the <located> and <subject> properties use a @literal value and the concept itself is not part of a controlled vocabulary, then the Concept ID has no meaning. However, it is still needed for the document to be valid NewsML-G2.

The QCode is created by generating a random string, or one that can be guaranteed unique within the scope of the Item, and appending this to the scheme alias “dummy” with the colon separator:

```

<assert literal="int014">
  <concept>
    <conceptId qcode="dummy: 091013121256" />
    <name>The Metropolitan Opera House</name>
  </concept>
</assert>

```

The scheme alias “dummy” resolves to a scheme URI hosted for this purpose by the IPTC. The scheme alias is part of the IPTC remote catalog:

```

<catalog ... >
  ...
  <scheme alias="dummy"
    uri="http://cv.iptc.org/newscodes/dummy/"
  />
  ...

```

If a provider wishes to use a different scheme alias with the scheme URI defined by the IPTC, they would need to create an entry in their catalog:

```

<catalog ... >
  ...
  <scheme alias="other"
    uri="http://cv.iptc.org/newscodes/dummy/"
  />
  ...

```

25.2.2 Using Assert to map from IIM or XMP

Many media organisations use the IPTC’s IIM standard (Information Interchange Model), particularly for pictures. IIM has been embedded into professional picture workflows because it was adopted by Adobe for the “IPTC Header” fields in Photoshop. This has been succeeded by Adobe XMP and IPTC Core for XMP, (See [Mapping Embedded Photo Metadata to NewsML-G2](#))

When migrating IIM or IPTC Core (XMP) metadata to NewsML-G2, <assert> offers a convenient method for conveying the contents of some fields which express the location shown by a picture using machine-readable or human-readable values.

For example, the location of the subject of a picture is conveyed using the NewsML-G2 <subject> property. However, <subject> cannot directly include detailed geographic information that may be contained in the embedded metadata. An Assert can be used to convey this information:

```

<contentMeta>
  ...
  <subject type="cpnat:geoArea" qcode="geocodes:abc" />
  ...
</contentMeta>
<assert qcode="geocodes:abc">
  <type qcode="cpnat:geoArea" />
  <geoAreaDetails>
    <position latitude="35.689962" longitude="139.691634"/>
  </geoAreaDetails>
</assert>

```

25.3 Inline Reference

The NewsML-G2 <inlineRef> element complements the <assert> property described above. Whereas <assert> carries supplementary information about concepts referred to in NewsML-G2 metadata, <inlineRef> performs a similar function for concepts found in the content of the document, such as is carried in inlineXML, and Label or Block type elements.

For example, the payload of a News Item might be text in XHTML; part of the text may refer to the Metropolitan Opera House. That portion of the text can be linked to information about “The Met”, by placing the supplementary information in an Inline Reference,

An <inlineRef> wrapper refers to tags with local identifiers (XML Schema IDREFS). The content associated with the Inline Reference must be tagged by an element that supports an attribute of type ID (not necessarily named “id”), examples include the XHTML element, and the NITF <org> element

The <inlineRef> element is only available at the Power Conformance Level (PCL) and is of Flex1PropType, with additional attributes of @idrefs, as noted above, and any members of the Quantify Attributes Group, which allows a provider to express, for example, the relevance of the supplementary information provided.

25.3.1 Quantify Attributes Group

Name	Datatype	Note
confidence	Int100	An integer from 0 to 100 expressing the provider’s confidence in the information provided. A value of 100 indicates the highest confidence
relevance	Int100	An integer from 0 to 100 expressing the relevance of the information provided. A value of 100 indicates the highest relevance.
why	QCode	A QCode from a provider-specific scheme indicating the reason that the information was included, for example that it was generated by software, or added by an editor.

25.3.2 Linking text content to an Inline Reference

A simple example illustrates the use of <inlineRef> to provide supplementary information about a concept found in an XHTML press release. In this case, the article refers to “the Met”. With 100% confidence, the provider declares that the name of this concept is “The Metropolitan Opera” (linking @id and @idrefs highlighted in **bold**):

```
<inlineRef idrefs=“xyz” confidence=“100” qcode=“acmeorg:int020”>
  <name role=“nrol:display”>Metropolitan Opera</name>
  <name xml:lang=“en-US” role=“nrol:full”>
    New York Metropolitan Opera
  </name>
</inlineRef>
<contentSet>
  <inlineXML contenttype=“application/xhtml+xml”>
    <html xmlns=“http://www.w3.org/1999/xhtml”>
      <head>
        <title>Free Tosca Open House Announced</title>
      </head>
      <body>
        <h1>New York Met Announces Free Tosca Open House</h1>
        <p>NEW YORK (Agencies) - On Thursday, September 17, the
          <span id=“xyz”>Met</span> will launch its fourth season of
          free Open Houses, with the final dress rehearsal of Luc
          Bondy’s new staging of Puccini’s opera, starring Karita
          Mattila and conducted by Music Director James Levine.
        </p>
        <p>Three thousand free tickets, limited to two per person,
          will be available beginning at noon on Sunday,
          September 13, at the Met box office only. The rehearsal
          starts at 11am on September 17, with doors opening at 10:30am.
        </p>
      </body>
    </html>
  </inlineXML>
</contentSet>
```

25.3.3 Using <inlineRef> and <assert> together

Inline Ref can provide basic information about a concept, but more detailed information can be expressed by linking the concept in the content, via <inlineRef>, to an <assert> element, using @literal or @qcode.

25.3.3.1 Example 1: simple story mark-up

The following shows how part of a story text section is associated via an inlineRef to a concept for President George W Bush:

```
<assert qcode="pers: gw. bush">
  <name>President George W. Bush</name>
  <type qcode="cptType: 9"/>
</assert>
<inlineRef idrefs="x1" qcode="pers: gw. bush" confidence="50"/>
<inlineXML contenttype="application/xhtml+xml" xml:lang="en" ... >
  <html xmlns="http://www.w3.org/1999/xhtml" ... >
    <head><title/></head>
    <body>
      <!-- Inline2 -->
      <p><span id="x1">President Bush</span> makes a speech about Iran
        today at 14:00.</p>
      ...
    </body>
  </html>
</inlineXML>
```

The content refers to "President Bush", and the inline reference links the text to the concept with a 50% confidence that the person referred to in the text is G. W. Bush, as the elder President George Bush could be the intended reference.

25.3.3.2 Example 2: complex story mark-up

The following expands on Example 1, providing further story text section mark-up, and indicating the associated inlineRefs and asserts to the concepts of President George Bush and Iran.

```
<assert qcode="pers: gw. bush">
  <name>President George W. Bush</name>
  <type qcode="cptType: person"/>
</assert>
<assert qcode="pers: g. bush">
  <name>President George Bush</name>
  <type qcode="cptType: person"/>
</assert>
<assert qcode="N2: IR">
  <name>Iran</name>
  <type qcode="cptType: geoArea"/>
</assert>
...
<inlineRef idrefs="x1" qcode="pers: gw. bush" confidence="50"/>
<inlineRef idrefs="x1 x3" qcode="pers: g. bush" confidence="70"/>
<inlineRef idrefs="x7" qcode="N2: IR" confidence="100"/>
...
<inlineXML contenttype="application/xhtml+xml" xml:lang="en" ... >
  <html xmlns="http://www.w3.org/1999/xhtml" ... >
    <head><title/></head>
    <body>
      <!-- Inline2 -->
      <p><span id="x1">President Bush</span> makes a speech about
        <span id="x7">Iran</span> today at 14:00.</p>
      <p>Later, <span id="x3">Bush</span> indicated that there were
        still many issues to be addressed.</p>
      ...
    </body>
  </html>
</inlineXML>
```

The mark-up associated with 'President Bush' implies:

- ❖ a 50% confidence in 'President George W. Bush' (via `inlineRef/@idrefs="x1"`), and
- ❖ a 70% confidence in 'President George Bush' (via `inlineRef/@idrefs="x1 x3"`).

The mark-up associated with 'Iran' implies:

- ❖ a 100% confidence in 'Iran' (via `inlineRef/@idrefs="x7"`), indicating it is a concept type of Geopolitical Unit (`cptType: geoArea`), refined as a Country (`geoProp: 5`).

25.3.3.3 Example 3: label mark-up

The following shows how a label is associated with an inlineRef (and therefore associated with an assert) as per Example 2:

```
<!-- contentMeta -->
<headline><span id="x8">President Bush</span> in
<span id="x7">Iran</span>. </headline>
...
<assert qcode="pers: gw. bush">
  <name>President George W. Bush</name>
  <type qcode="cptType: 9"/>
</assert>
<assert qcode="N2: IR">
  <name>Iran</name>
  <type qcode="cptType: 5" />
</assert>
...
<inlineRef idrefs="x8"      qcode="pers: gw. bush"   confidence="50"/>
<inlineRef idrefs="x7"      qcode="N2: IR"           confidence="100"/>
...
```

25.4 Why and How metadata has been added: @why and @how

If a provider needs to document the reasons for the presence of a particular metadata value, the @how and @why attributes are available for use with most elements as part of the Common Power Attributes (see Common Power Attributes table below).

The @why tells the end-user what caused the corresponding property to be added to the Item, for example that a person added the property based on their editorial judgement; @how tells the end-user the means by which the metadata value was extracted, for example that it was looked up from a database.

The properties are inter-dependent and used mutually exclusively: when the value of @why is "direct", meaning that the metadata was added by a person and/or a software tool directly from the content of the Item, it should be omitted and @how used to describe the method of extraction, as shown in the following code snippet illustrating its use with <subject>:

```
<subject
  type="cpnat: abstract"
  qcode="mysubject: JVN"
  <!-- @why omitted, as @how only used with @why value "direct" -->
  how="howextr: assisted">
  <name>
    JOINT VENTURES
  </name>
</subject>
```

If the value of @why is other than "direct", this implies that the value was extracted by a tool and therefore @how should be omitted because it would be redundant:

```
<subject
  type="cpnat: abstract"
  qcode="mysubject: JVN"
  <!-- @how omitted, as @why value "inferred" implicitly used a tool -->
  why="whypresent: inferred">
  <name>JOINT VENTURES</name>
</subject>
```

To promote interoperability the IPTC recommends that implementers use the IPTC NewsCodes How Extracted (<http://cv.iptc.org/newscodes/howextracted/>) with a Scheme Alias of "howextr" and Why Present (<http://cv.iptc.org/newscodes/whypresent/>) with a Scheme Alias "whypresent".

The Why Present CV contains three values:

- ❖ direct: the value is directly extracted from the content by a tool and/or by a person.
- ❖ ancestor: the value is an ancestor of another property value already assigned to the content (this implies that the ancestor was derived by a tool)
- ❖ inferred: the value was derived by look-up in a taxonomy/database (also implicitly by a tool).

The How Extracted CV also contains three values, used only with a @why value of "direct":

- ❖ assisted: the value was extracted by a person assisted by a tool.
- ❖ person: the value was extracted by a person.

❖ tool: the value was extracted by a tool.

25.5 Adding customer-specific metadata

In cases where customers have their own private metadata added by the provider of G2 Items, in order to facilitate processing and/or routing, the provider may not want to create specific versions of the G2 Item for each customer.

The solution is to use @custom, available to most elements (see details below) in conjunction with the attributes @how and @why to add the custom metadata.

Example 1: Use a customer's rule to express <memberOf> information

A news provider uses a customer's metadata rules to signal to the customer that the content is part of a collection of news entitled "Greek Debt". In the code snippet below, @creator expresses that the client rule was used to add this metadata, and @modified contains the date that the property was added. The @literal contains the customer's code for "Greek Debt" (also given in the name child element)

When @custom is set to "1" or "true" as in this example, this is a signal to receivers that the metadata may NOT exist in other instances of a G2 Item of the same guid and version number.

The @why attribute uses the "inferred" value from the IPTC Why Present NewsCodes, and this defines that the property was derived by lookup in a taxonomy or database. The @how attribute is not used as it is implicit that the property was extracted using a tool.

```
<memberOf
  literal="greek-dbt"
  creator="clientrule:r49"
  modified="2012-01-13T16:36:00Z"
  custom="true"
  why="whypresent:inferred">
  <name xml:lang="en">Greek Debt</name>
</memberOf>
```

25.5.1 Common Power Attributes group

The @custom attribute is one of several grouped under the Common Power Attributes, available to ALL elements in NewsML-G2 except the root elements of Items, elements of Ruby mark-up and a few elements of the News Message that are not root elements. The following table details the Group:

Definition	Name	Cardinality	Datatype	Notes
The local identifier of the element	id	(0..1)	XML Schema ID	
The person, organisation or system that created or edited the property.	creator	(0..1)	QCodeType	If the corresponding property has no value it specifies which party (person, organisation or system) will edit the property.
The date (and, optionally, the time) when the property was last modified. The initial value is the date (and, optionally, the time) of creation of the property.	modified	(0..1)	DateOptTimeType	



If set to true the corresponding property was added to the G2 Item for a specific customer or group of customers only.	custom	(0..1)	Boolean	The default value of this property is false which applies when this attribute is not used with the property.
Why the metadata has been included.	why	(0..1)	QCodeType	The recommended IPTC CV is http://cv.iptc.org/newsCodes/whypresent/ with a scheme alias of "whypresent"
The means by which the value was extracted from the content.	how	(0..1)	QCodeType	The recommended IPTC CV is http://cv.iptc.org/newsCodes/howextracted/ with a Scheme Alias of "howextr"
One or more constraints that apply to publishing the value of the property. Each constraint also applies to all descendant elements	pubconstraint	(0..1)	QCodeList	Use case: a publishable G2 document may contain metadata, or content references that are private, or restricted to certain audiences.

25.5.2 Aligning <subject> and <keyword> element properties, adding @how and @why

Since the <keyword> property was added to NewsML-G2 in v 2.4, it has become apparent that its use case aligns closely with the pre-existing <subject> element. In v 2.10, the attributes @confidence and @relevance were added to <keyword>, and @why and @how were added to both elements.

Both @confidence and @relevance are helpful to end-users who wish to filter results that are generated by software using linguistic processing and/or Bayesian inference.

Relevance is a measure of the usefulness of the concept to the user who wants to learn more about the principal subject matter of the content. A high relevance indicates that the metadata truly expresses the essence of the content, while a low relevance indicates a low correlation between the metadata and the content. For example, a concept that references Barack Obama would be 100 per-cent relevant to an article that about the 2012 U.S. presidential election. This would be represented as:

relevance="100"

A concept referencing The White House would be less relevant, and might be represented as:

relevance="60"

Confidence asserts the reliability of the association between the concept and the content and is usually obtained by software engines that automatically categorise news, where "100" is the highest value and "0" the lowest:

confidence="75"

Example 1: Keyword

In the code snippet, the first keyword was extracted directly from the content by a person with 100% relevance and confidence. The second keyword was "inferred" from the content (@how not present as it defaults to "tool") with 100 per cent relevance but 75 per cent confidence.

<contentMeta>


```

...
<keyword
  role="KeyRole: TOPIC"
  relevance="100"
  confidence="100"
  how="howextr: person">
    Armed Conflict
</keyword>
<keyword
  role="KeyRole: TOPIC"
  relevance="100"
  confidence="75"
  why="whypresent: inferred">
    Civil Unrest
</keyword>
...
</contentMeta>

```

Example 2: Subject

In this snippet taken from an announcement by a company, the subject property relating to the company is clearly highly relevant and could trigger the inclusion of further information about the company being included, for example in an <assert> wrapper. However, another company that is only mentioned in passing is only given a relevance of 50 per cent.

```

<contentMeta>
...
<subject
  type="cpnat: organisation"
  relevance="100"
  confidence="100"
  how="howextr: assisted"
  literal="More Widgets">
    <name>More Widgets Corp</name>
</subject>
<!-- Secondary company - relevance < 100 -->
<subject
  type="cpnat: organisation"
  relevance="50"
  confidence="100"
  how="howextr: assisted"
  literal="First Acme">
    <name>First Acme Mercantile Bank</name>
</subject>
...
</contentMeta>

```

Note that the value of @how in both cases is “assisted” and the confidence is 100 per cent, but the relevance is not a reflection of the method used to extract the metadata.

25.6 The derivation of metadata: the <derivedFrom> element

Although fixed taxonomies have their place in helping end-users classify news, there is common use of automation to “mine” news for concepts that can be used to create links and relationships. In such cases, it can be useful – in some cases essential – to create an “audit trail” of the derivation of metadata.

For example, if a journalist adds a subject code to an article, it is possible to express the reason that concept was added using the attributes @why and @how (see above)

In NewsML-G2 v 2.10, an attribute @derivedfrom was added to elements that supported the attribute @why. However, this restricted the expression of the derivation to property values expressed as QCodes: in fact the use case for derivation is often to give this attribution to metadata encoded outside the G2 framework.

In NewsML-G2 v 2.12, with an extension of @why to almost all G2 elements, it was decided to add a specific <derivedFrom> element to G2 as a child of the root element and using IDREFS to express the relationship between the derivedFrom element and the referring property¹⁶.

Note the direction of the relationship, as illustrated by the following snippet: an article about an event in Barcelona uses the event’s Twitter hashtag “#sgbcn” as a keyword. The editor of the article subsequently creates a G2 event with a Concept ID of “myevents:SGBCN” and indicates by the derivedFrom element

¹⁶ The attribute @derivedfrom is deprecated from NewsML-G2 v 2.11 on

that the <keyword> with the hashtag was *derived from* this concept. This is done by applying the keyword's id "s1" to the idrefs of the <derivedFrom> element which represents the event in Barcelona:

```
<newsItem>
...
  <contentMeta>
    <keyword id="s1" role="drol:hashtag">#sgbcn</keyword>
  </contentMeta>
  <derivedFrom idref="s1" qcode="myevents:SGBCN" />
...
</newsItem>
```

Based on this it can be asserted that the value of the keyword property with the id "s1" has been derivedFrom the event concept identified by the QCode "myevents:SGBCN".

25.7 Generic properties for proprietary use

NewsML-G2 implementers will sometimes encounter a use case that does not fit the defined semantics of an existing G2 metadata property. Using a property under an Extension Point is an option, since any user-defined structure may be added, but this requires the extension properties to be created in a different namespace. Receivers of proprietary properties may be unable to use a standard G2 property software module for processing such proprietary properties.

In NewsML-G2 v 2.17, a generic extension property was added to most elements wrapping a group of metadata properties. This enables a proprietary property to be expressed using a common type which includes @rel to define the relationship with the target concept. Even though the property may convey proprietary *data*, such generic properties can now be processed in a consistent manner due to their consistent structure. As a fictional example, the code below shows how a property could be defined to express a proprietary audience suitability rating for the content. The semantics of the extension property are defined using @rel, followed by the value of the property and its datatype:

```
<contentMeta>
...
  <contentMetaExtProperty
    rel="mediarel:suitableAudience"
    value="PG"
    valueType="xs:string"
  />
...
</contentMeta>
```

Where there are time restrictions on the use of the extension property, these may be expressed using @validfrom and @validto.

25.7.1 Full list of available extension properties

[Should the entries in col 1 and/or 3 be hyperlinks?]

Wrapper	Extension Property	Affected Item(s)
rightsInfo	rightsInfoExtProperty	all
itemMeta	itemMetaExtProperty	all
contentMeta	contentMetaExtProperty	All except catalogItem
partMeta	partMetaExtProperty	All except catalogItem
remoteContent	remoteContentExtProperty	newsItem
concept	conceptExtProperty	conceptItem, knowledgeItem
group	groupExtProperty	packageItem
newsCoverage	newsCoverageExtProperty	planningItem
planning	planningExtProperty	planningItem
header	headerExtProperty	newsMessage
schemeMeta	schemeMetaExtProperty	knowledgeItem

25.8 Special processing instructions for metadata about a part of the content



Implementers need to be aware of how property values about the content as a whole are inherited by individual parts of the content.. All property assertions in elements such as `<itemMeta>` or `<contentMeta>` are inherited by any `<partMeta>` element and are specifically over-ridden for that part of the content when re-stated in a `partMeta` element.

An example of such inheritance is the use of the `<signal>` element as a child of `<itemMeta>` and/or `<partMeta>`:

The `<signal>` element instructs the G2 processor to handle an Item or its content in a specific way. When used as a child element of `<itemMeta>` its scope is the entire Item; when used as a child of `<remoteContent>` its scope is the specific rendition of the content. By introducing `<signal>` at the `<partMeta>` level, providers can specify processing instructions for any part of the content.

When used in a `<partMeta>` element, any processing instructions expressed at a higher level in the item are over-ridden for that specific part of the content. Thus, using `<signal>` as an example processing instruction:

```
<itemMeta....>
  <!-- A applies to the whole item, unless overridden -->
  <signal... A/>
</itemMeta>
<partMeta id="1">
  <!-- A does NOT apply to this content part -->
  <!-- B applies to this content part only -->
  <signal... B/>
</partMeta>
<partMeta id="2">
  <!-- inherits only the signal A from Item Metadata-->
  <!-- B does NOT apply to this content part -->
</partMeta>
```

Where a processing instruction (from a higher level) is NOT to be inherited for a specific part of the content, that processing instruction needs to be explicitly "nulled" within that specific part of the content. Using `<signal>` again as an example processing instruction, a "null" value is added to the content part:

```
<partMeta id="3">
  <signal qcode="gen: null" /> <!-- nulls inheritance of any signal(s) from the Item
Metadata level -->
</partMeta>
```



26 Generic Processes and Conventions

This Chapter discusses some processes, procedures and conventions which are generic to all G2 Items, and relate to best practice in the wider context of news processing.

26.1 Processing Rules for NewsML-G2 Items

Although at CCL, NewsML-G2 is designed for maximum inter-operability, providers are strongly recommended to document their implementation of NewsML-G2 and the provider-specific rules and conventions used.

This information must be provided to receivers so that generic NewsML-G2 processors can be parameterised accordingly

An obvious example is Packages (see [Package Processing Considerations](#)) where the structure and content types must be pre-arranged between provider and receivers in order to facilitate correct processing.

Another example is the use of property attributes such as @role and @type which refine the semantic of the property. There needs to be a clear understanding of the concepts being used in these circumstances for the full value of the information to be useful.

26.2 Using Links

26.2.1 Introduction

The <link> property is used in the <itemMeta> section of a NewsML-G2 document to create a navigable link between the Item and a related resource. Examples of the target of a link could be a Web page, a discrete object such as an image file, or another NewsML-G2 Item.

Valid uses of <link> include:

- ❖ To indicate a supplementary resource, for example a picture of a person mentioned in an article.
- ❖ To identify the resource that an Item is derived from, for example if providing a translation.
- ❖ where systems do not support versioning, to provide a link to the previous version of an Item.
- ❖ To identify the previous “take” of a multi-page article, or the previous Item in a series of Items. (Note that these use cases are not explicitly supported by the current version of the IPTC Item Relationship NewsCodes. The <memberOf> property of Item Metadata expresses that an Item is part of a series of Items, but not its sequence.)
- ❖ Where a News Item conveys formatted text which references an illustration, a dependency link from the article to the illustration is indicated using <link>.

26.2.2 Link properties

Link uses the LinkType datatype (CCL), with optional attributes for Item Relation (@rel) and the Target Resource Attributes group. For each <link> at CCL, any number of child <title> properties of the target resource may be added.

At PCL the Link1Type datatype is used, optionally with more extensive attributes, which permits any property consistent with the structure of the target resource to be used as a processing hint. See [26.2](#).

26.2.2.1 Item Relation @rel and the Item Relationship NewsCodes

A QCode indicating the relationship of the current Item to the target resource. For example, if the current Item is a translation from an original article, the relationship may be indicated using the IPTC Item Relation NewsCode, for example:

```
<link rel="irel:translatedFrom">
```

The default relationship between the host Item and a resource identified by a <link> is “See also”. The CV broadly defines three types of relationship:



- ❖ Navigation: “See Also”.
- ❖ Dependency: “Depends On”.
- ❖ Derivation: includes “Derived From” and other refinements of derivation relationships.

The IPTC recommends that for derivation relationships, implementers should use the most specific available representation. For example if a picture conveyed by the item is a crop of the image indicated by <link>, “Derived From” is not inaccurate, but “Cropped From” is preferred as it is more specific.

26.2.2.1.1 IPTC Item Relationship CV (Scheme URI <http://cv.iptc.org/newscodes/itemrelation/>)

Relationship	Code	Definition
See Also	seeAlso	The related item or resource can be used as additional information. Full rendering of this Item does NOT depend on the retrieval of the related item or resource
Depends On	dependsOn	Full rendering of this Item depends on the retrieval of the related item or resource.
Derived From	derivedFrom	This item was derived from the related item or resource.
Associated With	associatedWith	RETIRED
Previous Version	previousVersion	The related item is the previously published version of this item.
Translated From	translatedFrom	This item is a translation of the related item or resource
Cropped From	croppedFrom	The visual content of this item has been cropped from the related item or resource.
Evolved From	evolvedFrom	This item was evolved from the related item. Typical use is for stories evolving over time, including corrections.
Processed From	processedFrom	The content of this item has been processed from the content of the related item or resource. Do not use in the case of cropping (see croppedFrom).
Previous Version	prevVersion	RETIRED. Replaced by previousVersion (see above)

26.2.2.2 Target Resource Attributes

These are property attributes that enable the receiver to accurately identify the remote resource, its content type and size See [9.7.2 Target Resource Attributes](#)

26.2.2.3 Item Title <title>

A short natural language name describing the link, intended to be displayed to users. It is not necessary that this <title> should be extracted from the target resource. For example, a journalist may wish to add a link and write a title for it.

`<title xml:lang="en-GB">File picture of President Obama</title>`

26.2.2.4 Hint and Extension Point

At PCL any number of metadata properties from the IPTC’s News Architecture (NAR) namespace may be included. For example, a picture referenced by a link may have the recommended filename extracted from the target resource as an aid to processing. The inclusion of hints is subject to the following:

- ❖ The properties MUST comply with the structure of the NewsML-G2 Item target resource but do NOT have to be extracted from it. (In other words, they could exist in the target resource but it is not mandatory that they do so),
- ❖ Immediate child elements of <itemMeta> or <contentMeta> and <concept> optionally with their descendants, may be used directly.
- ❖ For all other elements, the full path must be used, excluding the root element of the NewsML-G2 Item.



The rules are to avoid confusion when using properties that could be used simultaneously in different structures with the same Item, for example <description> is a child of both <contentMeta> and <partMeta>. These rules enable implementers to place metadata in the correct context. An example is given in **12.7 Relationships between Concepts**

26.2.3 Link examples

26.2.3.1 A supplemental picture with a text article (CCL)

The sender of a News Item containing a text article wishes to include a link to a picture that may optionally be retrieved to illustrate the article. The relationship to the target resource is “see also”.

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:acmenews.com,2008:TX-PAR:20090529:JYC85"
  version="1"
  ....>
  <itemMeta>
    <itemClass qcode="ni nat: text" />
    ....
    <edNote>With picture </edNote>
    <link
      rel="i rel: seeAlso"
      residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC80">
      <title>
        File picture of President Obama
      </title>
    </link>
    ....
  </itemMeta>
  ....
</newsItem>
```

← Item relation
← Item ref
← Title

26.2.3.2 A required picture with a text article (PCL)

A News Item contains a text article which is marked up explicitly to reference a picture (for example in XHTML). The picture is required for correct display of the text-with-picture story, therefore the relationship to the target resource is “depends on”.

A NewsML-G2 processor should be enabled to pre-fetch this required target resource before the content of the news item is processed. As shown in the example below, the “real world” link between the article and the picture is established using the filename of the linked resource.

This example is at PCL; we are able to extract the <itemClass> property and <filename>, the recommended filename of the target resource (a News Item), from the Item’s <itemMeta> wrapper and add these as child elements of <link>.

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:acmenews.com,2008:TX-PAR:20090529:JYC85"
  version="1"
  standard="NewsML-G2" standardversion="2.18"
  conformance="power"
  ....>
  <itemMeta>
    <itemClass qcode="ni nat: text" />
    ....
    <edNote>With picture </edNote>
    <link
      rel="i rel: dependsOn"
      residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC80">
      <itemClass qcode="ni nat: text" />
      <filename>
        obama- omaha- 20090606. jpg
      </filename>
    </link>
    ....
  </itemMeta>
  <contentSet>
    <inlineXML>
      <html xmlns="http://www.w3.org/1999/xhtml">
        ....
        <p>At Omaha Beach, President Obama led a ceremony to mark the landing of
          thousands of U.S. troops on D-Day.</p>
      </html>
    </inlineXML>
  </contentSet>
  ....
</newsItem>
```

← Item relation
← Item ref
File name

```
  
...  
</inlineXML>  
</contentSet>  
</newsItem>
```

26.2.3.3 Linking to previous versions of an Item (PCL)

(See also **Processing Updates and Corrections**) Some content management systems do not maintain a common identifier for successive versions of an information asset (text, picture etc), but maintain a link to the identifier of the previous version of the asset. In these circumstances, a `<link>` can inform recipients that the current Item is a new version of a previously published Item, and provide a navigation to retrieve the previous version of the Item, if required.¹⁷ The relationship between the current item and the previous version is "previousVersion".

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"  
guid="tag:acmenews.com,2008:TX-PAR:20090529:JYC85" ← new item  
version="1" ← first version  
...>  
<itemMeta>  
<itemClass qcode="ni nat: text" />  
...  
<edNote>Replaces previous version. MUST correction, updates name of  
minister</edNote>  
<signal qcode="action: replace" />  
<link  
rel="i rel: previousVersion"  
residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC80" ← previous item  
version="1" > ← first version  
<itemClass qcode="ni nat: text" />  
</link>  
...  
</itemMeta>  
...  
</newsItem>
```

Using this method, other target resources such as previous "takes" of a multi-page article, the original picture from which the current item was cropped, or the original text from which a translation has been made, can be expressed using `<link>`

26.3 Publishing Status

The NewsML-G2 `<pubStatus>` property uses a mandatory IPTC CV that contains three values:

- ❖ usable
- ❖ withheld
- ❖ canceled

These terms have a specific meaning in a professional news workflow, and it is the IPTC's intention in designing NewsML-G2 that they be interpreted by software systems. They are NOT intended as advisory notices to journalists, although of course the Publishing Status may well be a read-only property displayed by an editing system.

If no `<pubStatus>` property is present in an Item, the default value is "usable", meaning that the item and its contents may be published.

If an item has a publishing status of "withheld", this signals that the item and its contents may NOT be published until further notice. That status may be published only after receipt of a new version of the item – using the same GUID – that has a status of "usable".

¹⁷ Even where providers use the same ID, it is not mandatory to use a consecutive ascending sequence of numbers to indicate successive versions of an Item. Where it is required to positively identify the previous version of an Item, a provider SHOULD add a `<link>` to the previous version.

For example, a provider may send an item of news (version 1), and subsequently decided that a correction or amplification is needed, which requires the sending of a new version of the item. If the new version will not be ready for an appreciable time, the provider may send a new version (version 2) of the item with a status of “withheld” to stop further publication of the incorrect item. When the corrected version is ready, it will be sent – using the same GUID – with a status of “usable” (version 3).

An item with a status of “withheld” **MUST NOT** be published. It may only have its status changed to “usable”, at which point it may be published, or “canceled”.

If an error cannot be corrected, or the item needs to be permanently withdrawn for some other reason, the provider may use “canceled” the third value of <pubStatus> (note U.S. spelling). This instructs receiving systems to remove all versions of the item from all locations, including (and especially) archives. News organisations have faced legal action arising from the inadvertent re-publication from an archive of defamatory content.

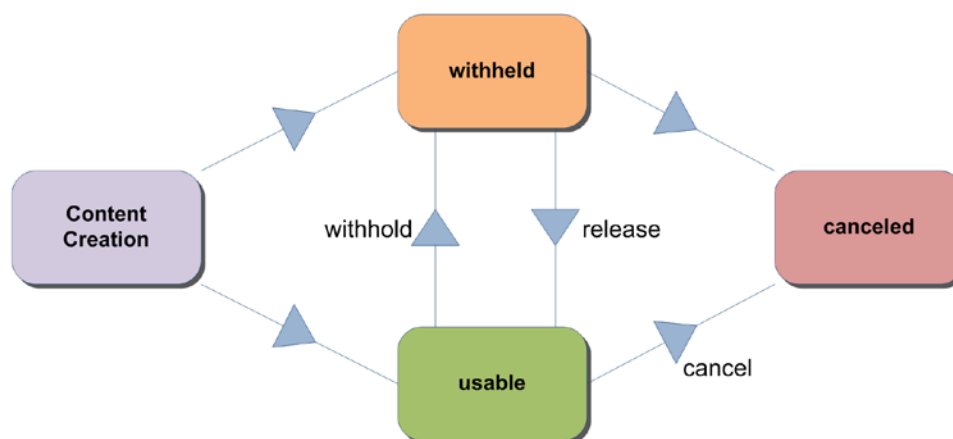


Figure 23: State Transition Diagram for <pubStatus>

A “canceled” item **CANNOT** have its status changed back to “usable” or “withheld”. If a provider wishes to send revised content, it **MUST** be sent under a NEW GUID.

The <pubStatus> property is part of the <itemMeta> component, and uses a QCode value. The scheme alias for the IPTC Publishing Status NewsCode is “stat”:

```
<pubStatus qcode="stat:usable" />
```

26.4 Embargo

Professional, or business-to-business, news organisations often make use of an *embargo* to release information in advance, on the strict understanding that it may not be released into the public domain until after the *embargo time* has expired, or until some other form of permission has been given.

Embargo is **NOT** the same as the Publishing Status. Some systems process the embargo time using software in order to trigger the release of content when the embargo time is passed, but the intention of embargo is also as an information management feature for journalists.

Embargos are generally an unwritten agreement and have no legal force. Their success depends on cooperation between parties not to abuse the system. Possible abuses include imposing unnecessary embargos in order to manage the impact of news, or by breaking embargos and releasing news into the public domain too early.

NewsML-G2 uses the optional <embargoed> property in <itemMeta> to indicate whether an item is under an embargo. If the property is absent there is no embargo:

```
<embargoed>2014-10-23T12:00:00Z</embargoed>
```

If the property is present **AND** empty, this enables providers to release an item under embargo when the precise date and time that the embargo expires is not known. In these circumstances, an <edNote> or

some contractual agreement between the provider and customer will specify the conditions under which the embargo may be lifted.

For example, a provider may release an advance copy of a speech which may not be released to the public until the speaker has finished delivering it. The provider would have no way of knowing exactly when this would be. Therefore some other means of authorising the release may be negotiated between the parties, such as email or a phone call:

```
<embargoed />
<edNote>
  Note to editors: STRICTLY EMBARGOED. Not for release until authorised. Our
  News Desk will advise your duty editor by email. Release expected about
  12noon on Monday, February 9.
</edNote>
```

Check the corresponding NewsML-G2 Specification Document for further information regarding processing rules for `<embargoed>`. The processing model is illustrated in [Figure 24](#).

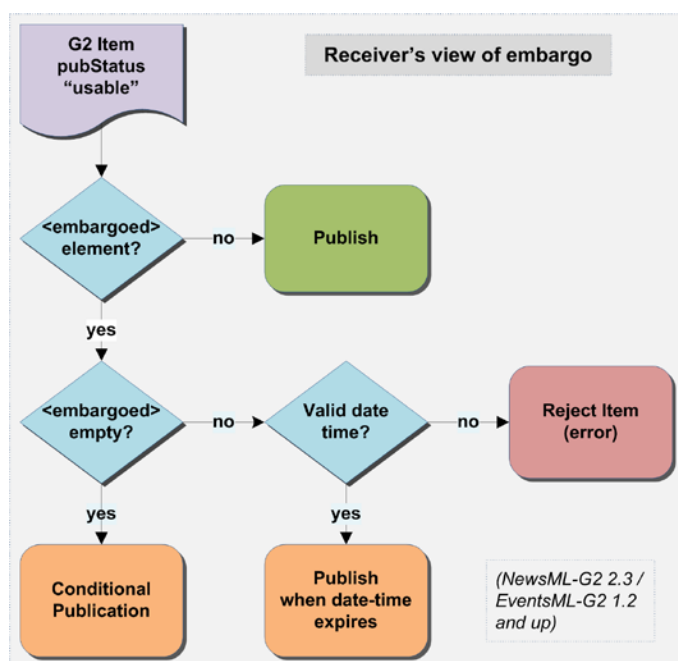


Figure 24: Processing `<embargoed>`

26.5 Geographical Location

There are two properties of `<contentMeta>` that express the geographic location(s) associated with a NewsML-G2 Item, with a distinction between their uses.

The Located `<located>` property expresses the origin of the content: where the content was created, for example a text article written or a picture taken. The intention of `<located>` is as a machine-readable equivalent to the location given in a Dateline. (The NewsML-G2 `<dateline>` property is also available as a natural-language string, for example "MOSCOW, Monday (Reuters)")

The accepted convention, which in some news organisations is formalised as part of a code of practice, is that the Dateline identifies the place where content is created, NOT the place where an event takes place. They may be the same, but this is not necessarily the case.

- ❖ In conflict zones, journalists may not have access to the area where reported events are taking place.
- ❖ Even when physical access is not an issue, journalists may have relied on interviewing people by telephone or on reports from freelance correspondents in order to get the material used to write an article.



The Located property is therefore provided to express the place of origin of content as part of Administrative Metadata:

```
<located type="locotype: city" qcode="city: 75000">
  <name>Paris</name>
</located>
```

To express the geographical information that is important in the context of the article or picture, the Subject property <subject> is used, optionally using a Concept type (@type). In the example below "cpnat:geoArea" from the IPTC Concept Nature NewsCode, is used, but providers may have their own scheme(s).

The Subject property is Descriptive Metadata:

```
<subject type="cpnat:geoArea" qcode="city: Tblisi">
  <broadener qcode="cntry: Georgia" >
    <name>Georgia</name>
  </broadener>
</subject>
```

This news story fragment from Reuters and the accompanying code listing illustrate the use of <located> and <subject>. The geographical subjects of the report are Georgia and South Ossetia, but the report was written in Moscow:

MOSCOW, Monday (Reuters) - The breakaway Georgian region of South Ossetia alleged today that two unexploded Georgian shells landed in its capital Tskhinvali, but Tbilisi dismissed the claim as nonsense.

Both sides have regularly....

LISTING 28 Illustrating Located, Subject and Dateline

```
<contentMeta>
  <contentCreated>2007-02-09T09:17:00+03:00
  </contentCreated>
  <located qcode="city: Moscow">
    <broadener qcode="cntry: Russia" />
  </located>
  <creator qcode="web: thomsonreuters.com">
    <name>Thomson Reuters</name>
  </creator>
  <language tag="en-US" />
  <subject type="cpnat:geoArea" qcode="city: Tskhinvali">
    <broadener qcode="reg: SouthOssetia" >
      <name>South Ossetia</name>
    </broadener>
  </subject>
  <subject type="cpnat:geoArea" qcode="city: Tblisi">
    <broadener qcode="cntry: Georgia" >
      <name>Georgia</name>
    </broadener>
  </subject>
  <dateline>MOSCOW, Monday (Reuters)</dateline>
</contentMeta>
```

26.6 Processing Updates and Corrections

By its nature, news may need frequent updating, and in some cases correcting, as new facts emerge. The simplest NewsML-G2 mechanism for dealing with updated content is to re-issue an item using the same GUID with a new Version.

In the absence of any specific instructions from the provider, a "usable" item should be regarded as replacing any previous version of the item with the same GUID. In practice, a provider is likely to provide some supplementary information in the form of a human-readable <edNote> which can be displayed to inform recipients of the reason for the update.



NewsML-G2 also provides machine-readable features to express whether a new version updates or corrects previous versions of an Item and a further indication of the impact of the change, using the <signal> property under Item Metadata at CCL and PCL.

Signal uses a QCode to identify an action from a CV. To promote interoperability, the IPTC maintains NewsCodes for this purpose, but note it is up to the provider to specify the rules for applying the codes so that their end-users can correctly process the instruction.

26.6.1 Signalling Update or Correction

Any version of an item except the initial version is implicitly an update of the previous version. It is not required to use the update signal, but it is not always possible to infer from the version number whether a document is an initial version or is implicitly an update. Therefore the IPTC recommends that <signal> is used with the IPTC Signal NewsCode (Scheme URI <http://cv.iptc.org/newscodes/signal/> and recommended Schema Alias "sig"). The relevant members of the Scheme are:

Code	Definition	Note
update	This version of the Item includes an update of some part of the previous version of the Item.	Implies that this version of the Item is not the initial version
correction	This version of the Item corrects some part of a previous version of the Item.	Implies that this version of the Item is not the initial version. This Correction signal does not indicate in which version(s) of the Item the corrected error existed.

In addition, the Editorial Note <edNote> property under Item Metadata may be used to provide natural language details about the update or correction, such as specifying a name in the text that has been corrected or whether a paragraph with updated information has been added to the text.

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:afp.com,2008:TX-PAR:20090529:JYC80"
  version="2"
  ...>
  <itemMeta>
    ...
    <edNote>Updates previous version by appending these paragraphs</edNote>
    <signal qcode="sig:update" />
    ...
  </itemMeta>
</newsItem>
```

← current item new version

26.6.2 Signalling the impact and reason for an update

Whether or not the update signal is used, a concept from the Update NewsCodes (<http://cv.iptc.org/newscodes/update/>) may be used as a refinement of the reason for distributing an Item update.

The three axes of update that are combined to inform receivers how the new version may be processed are:

- ❖ Operation: Replace or Append
- ❖ Impact: Major or Minor change
- ❖ Parts affected: Metadata or Content or Both (default is "Both")

This results in CV values as follows:

Code	Definition
replace-minor	This item replaces the previous version of the Item with minor changes to both content and metadata.
replace-minor-metadata	This item replaces the metadata in the previous version of the Item with minor changes.
replace-minor-content	This item replaces the content in the previous version of the Item with minor changes
replace-major	This item replaces the previous version of the Item with major changes to both content and metadata.

Code	Definition
replace-major-metadata	This item replaces the metadata in the previous version of the Item with major changes.
replace-major-content	This item replaces the content in the previous version of the Item with major changes
append-minor	This item contains minor additions to both content and metadata that should be appended to the previous version of the Item
append-minor-metadata	This item contains minor additions to the metadata that should be appended to the metadata in the previous version of the Item
append-minor-content	This item contains minor additions to the content that should be appended to the content in the previous version of the Item
append-major	This item contains major additions to both content and metadata that should be appended to the previous version of the Item
append-major-metadata	This item contains major additions to the metadata that should be appended to the metadata in the previous version of the Item
append-major-content	This item contains major additions to the content that should be appended to the content in the previous version of the Item



The “append” codes are a special case that requires a specific processing rule imparted by the provider to customers. If the receivers are being instructed to perform the “append” operation, then they need to know precisely on which version of a document to perform the operation, and this may not be clear from the version number alone (it is not mandatory to start versioning at “1” or use consecutive version numbers).

If using the “append” NewsCodes with signal, providers SHOULD add a <link> to the previous version of the document.

As before, the Editorial Note <edNote> property under Item Metadata may be used to provide natural language details about the update. For example:

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:afp.com,2008:TX-PAR:20090529:JYC80"
  version="234256"
  ...>
  <itemMeta>
    ...
    <link
      guidref="tag:afp.com,2008:TX-PAR:20090529:JYC80"
      version="1234" />
    <edNote>URGENT: Append judge's statement to previous version </edNote>
    <signal qcode="sig:update" />
    <signal qcode="update:append-major-content" />
    ...
  </itemMeta>
</newsItem>
```

← current item
new version

26.7 Content Warning

The <signal> property in Item Metadata may be used to inform end-users that the nature of the content being sent may be perceived as being offensive to some audiences. This uses the IPTC Signal NewsCode, with Scheme URI <http://cv.iptc.org/newscodes/signal> and recommended Schema Alias “sig” with a code of “cwarn”:

```
<signal qcode="sig:warn" />
```

The Signal property is available at CCL and PCL.

Optionally, the nature of the warning can be expressed using the Exclude Audience property <exclAudience> and using the IPTC Content Warning NewsCodes. The Scheme URI is <http://cv.iptc.org/newscodes/contentwarning/> and recommended Schema Alias is “cwarn”. The Scheme values are:

Code	Definition
death	The content could be perceived as offensive due to the discussion or

Code	Definition
	display of death.
language	The content could be perceived as offensive due to the language used..
nudity	The content could be perceived as offensive due to nudity.
sexuality	The content could be perceived as offensive due to the discussion or display of sexuality.
violence	The content could be perceived as offensive due to the discussion or display of violence.

For example:

```
<signal qcode="sig: warn" />
<exclAudience qcode="cwarn: nudity"/>
<exclAudience qcode="cwarn: language"/>
```

26.8 Working with Social Media

26.8.1 Ratings

The ability for end-users to rate or interact with content has undergone enormous growth as part of social media and the “socialising” of the Web, and has led to a clear business need that user actions and ratings must be expressed as part of the metadata for all kinds of content.

G2 provides a set of properties for implementers for a flexible use in expressing actions and ratings:

- ❖ The <userInteraction> element can be used to express interactions with the content of this item such as Facebook “likes”, tweets, and page views.
- ❖ The <rating> element contains means to express a rating value that applies to the content of this item, such as a star rating, and includes the ability to convey how many “raters” led to the rating, plus information about how the rating system is constructed and how the rating value was calculated.

Both elements are children of <contentMeta> in any type of G2 Item, and of <partMeta> in any G2 Item except for Concept Items and Planning Items.

26.8.1.1 The <userInteraction> element

This element expresses the number of times that end-users have interacted with content in the form of “Likes”, votes, tweets and other metrics. If the element is used, it has two mandatory attributes:

- ❖ interactions: the count of executed interactions (expressed as a non-negative integer)
- ❖ interactiontype: a QCode indicating the type of interaction which is reflected by this property. The proposed IPTC CV is User Interaction Type, Scheme URI <http://cv.iptc.org/newscodes/useractiontype/> with a recommended Scheme Alias of “useracttype”

The example below expresses the number of Facebook “Likes”, Google+ “+1” and tweets garnered by the content conveyed in the G2 document.

```
<contentMeta>
...
<userInteraction
  interactions="36"
  interactiontype="useracttype: fblike" />
<userInteraction
  interactions="22"
  interactiontype="useracttype: googleplus" />
<userInteraction
  interactions="1203"
  interactiontype="useracttype: tweets" />
...
</contentMeta>
```

...

The User Action Type NewsCodes have a Scheme URI of <http://cv.iptc.org/newscodes/useractiontype/> and a recommended Scheme Alias of "useracttype". NewsCodes in the scheme are:

NewsCode	Name	Definition
fblike	Facebook Like	Indicates that the user interaction was measured by Facebook's I Like feature.
googleplus	Google +1	Indicates that the user interaction was measured by Google's +1 feature.
retweets	Twitter re-tweets	Indicates that the user interaction was measured by re-tweets of a Twitter tweet.
tweets	Twitter tweets	Indicates that the user interaction was measured by tweets which mention the subject of the content.
pageviews	Page views	Indicates that the user interaction was measured by the number of page views of the content.

26.8.1.2 The <rating> element

Ratings such as "five stars" have existed for many years, in the news domain they have been a long-standing feature of image management software. Now they are used on the Web for all kinds of content as providers seek to engage audiences and promote feedback. However, different types of rating are needed – for example content may be rated in terms of how "useful" it was to the user – so the G2 <rating> element is a wrapper for attributes that can define different rating models, including :

- ❖ How the rating is expressed
- ❖ How many individual ratings exist
- ❖ How the rating was calculated

For ease of reading, the full attributes of <rating> are tabulated below:

Definition	Name	Cardinality	Datatype	Notes
The rating of the content expressed as decimal value	value	(1)	Decimal	
How the value is calculated such as median, sum	valcalctype	(0..1)	QCodeType	Proposed IPTC CV (see below).
Minimum rating value.	scalemin	(1)	Decimal;	
Maximum rating value.	scalemax	(1)	Decimal;	
Rating units	scaleunit	(1)	QCodeType	Two values in the proposed IPTC CV are "star" and "numeric".(see below)
The number of parties who contributed a rating.	raters	(0..1)	NonNegativeInteger	If not present, defaults to 1.



The type of the rating parties.	ratertype	(0..1)	QCodeType	Proposed IPTC CV (see below).
Full definition of the rating.	ratingtype	(0..1)	QCodeType	The rating engine or method used, for example Amazon star rating or Survey Monkey.

The following example expresses a star rating with a minimum rating of one star and a maximum of 5. The rating is 4.5 (the provider and end-user may need to agree on how to handle values that are not a whole star, for example by rounding and/or using half-stars). The rating was calculated from the arithmetic mean of ratings by 123 users.

```

...
<contentMeta>
...
  <rating
    value="4.5"
    valcalctype="rcalctype:amean"
    scalemin="1"
    scalemax="5"
    scaleunit="rscaleunit:star"
    raters="123"
    ratertype="rtype:user" />
...
</contentMeta>
...

```

26.8.2 Content Rating NewsCodes

As an aid to interoperability, the IPTC has created Controlled Vocabularies (NewsCodes) for the properties @valcalctype, @scaleunit and @ratertype:

26.8.2.1 Rating Calculation Type NewsCodes

These NewsCodes indicate how the applied numeric rating value was calculated from a sample of values. The Scheme URI is <http://cv.iptc.org/newscodes/rcalctype/> and the recommended Scheme Alias is "rcalctype". NewsCodes in the scheme are:

NewsCode	Name	Definition
amean	arithmetic mean	Indicates that the numeric value was calculated by applying the arithmetic mean algorithm to a sample.
gmean	geometric mean	Indicates that the numeric value was calculated by applying the geometric mean algorithm to a sample.
hmean	harmonic mean	Indicates that the numeric value was calculated by applying the harmonic mean algorithm to a sample.
median	median	Indicates that the numeric value is the middle value in a sorted list of sample values.
sum	sum	Indicates that the numeric value was calculated as the sum of a sample.
unknown	unknown	Indicates that the algorithm for calculating the numeric value from a sample is not known.

26.8.2.2 Rating Scale Unit NewsCodes

These NewsCodes indicate the units used to express the rating value. The Scheme URI is <http://cv.iptc.org/newscodes/rscaleunit/> and the recommended Scheme Alias is "rscaleunit". NewsCodes in the scheme are:

NewsCode	Name	Definition
star	star	Star rating
number	number	Numeric rating

26.8.2.3 Rater Type NewsCodes

These indicate the type of the party or parties that have contributed to a rating. The Scheme URI is <http://cv.iptc.org/newscodes/ratertype/> and the recommended Scheme Alias is "rtype". NewsCodes in the Scheme are:

NewsCode	Name	Definition
expert	expert	The persons who rated are experts in what they rated.
user	user	The persons who rated are using what they rated.
customer	customer	The persons who rated have bought and are using what they rated.
mixed	mixed	The persons who rated are not of the same rater type.

26.8.3 Hashtags

As these are essentially uncontrolled values, at least within the scope of NewsML-G2, so the recommended way to add these social media tags is by using the <keyword> property with a @role to indicate their purpose. A new NewsCode has been added to the IPTC Description Role scheme to enable implementers to do this in a standard way:

NewsCode	Name	Definition
hashtag	Hashtag	A word or phrase (with no spaces) which is prefixed with the hash character "#"

The Scheme URI is <http://cv.iptc.org/newscodes/descriptionrole/> and the recommended Scheme Alias is "drol", thus:

```
<contentMeta>
  ...
  <keyword role="drol:hashtag">#iptcrocks</keyword>
  ...
</contentMeta>
```

26.9 Indicating that a News Item has specific content

In some workflows, the content of a News Item may not be available at the time of the distribution of its first version. For example, if an organisation normally provides a video in HD and SD, but the HD rendition is not yet available, there is a lightweight way to inform end-users that the content is planned, using @hascontent.

This optional attribute may be added to any child element of <contentSet> (inlineXML, inlineData, remoteContent) as a flag to indicate whether the element contains content/content reference or is either



empty/only metadata. @hascontent is simply a Boolean flag with a default of "1" or "true", and "0" or "false" to indicate that this element contains no content.

The example below is a <remoteContent> showing that an HD rendition of the existing SD rendition of a video will follow soon. To make the receiver aware of the planned release of the additional rendition, the wrapper has been added to contentSet with only metadata defined, and no actual content being provided, and a @hascontent value of "isfalse":

```
<contentSet>
  <remoteContent
    href="http://www.example.com/video/2008-12-22/20081222-PNN-1517-
    407624/20081222-PNN-1517-407624.avi"
    format="fmt:avi"
    duration="111" durationunit="timeunit:seconds"
    videocharacteristic="videodef:SD"
    videoframerate="25"
    videoaspectratio="16:9" />
  <remoteContent
    hascontent="false"
    format="fmt:avi"
    duration="111" durationunit="timeunit:seconds"
    videocharacteristic="videodef:HD"
    videoframerate="25"
    videoaspectratio="16:9" />
</contentSet>
```

27 Identifying Sources and Workflow Actors

27.1 Introduction

NewsML-G2 contains a number of features designed to help providers and end-users maintain an audit trail for an Item, its payload and metadata, and in some cases to maintain document/content security.

27.2 Information Source <infoSource>

The <infoSource> property, together with its @role, enables finely-grained identification of the various parties who provided information used to create and develop an item of news. The definition of <infoSource> was extended in NewsML-G2 v2.10. It had previously been defined as the party that originated or enhanced the content. This was extended to "A party (person or organisation) which originated, modified, enhanced, distributed, aggregated or supplied the content or provided some information used to create or enhance the content."

This extension required that the @role attribute of Information Source be more rigorously defined in order to maintain inter-operability between providers and receivers, and promote a better understanding of the part played by the various actors identified in multiple <infoSource> elements.

The value of @role should be taken from the recommended IPTC Info Source Role NewsCode created and maintained for this purpose. The Scheme URI is <http://cv.iptc.org/newscodes/infosourcerole/> and the recommended Scheme Alias is "isrole".

The default value of @role is "isrole:originfo", indicating that the party identified originated information used to create or enhance the content, and may be omitted. If a party did anything more than originate information, one or more @role attributes MUST be applied. The table below shows the Info Source Role CV values and definitions..

Code	Name	Definition
interviewee	Interviewee	A person who granted an interview
mediaOffice	Media Office	The media relationships office of a person, organisation or company.
origcont	Content Originator	A party which originated the content of this item.
originfo	information originator	A party which provided some information used to create or enhance the content of this item.
spokesperson	Spokesperson	The spokesperson of a person, organisation or company

27.3 Identifying workflow "actors" – a best practice guide

The <infoSource> property is one of a set of properties in NewsML-G2 that are used to identify the actors in the creation and evolution of an item of news. The following table is a guide to how these properties should be used:

Property	Location in an Item	Usage notes
News Message		
Message Sender	header/sender	Party sending the <i>items</i> in the message, which may be an organisation or a person. Best practice is to identify a sender by a domain name.
Item		
Hop History	hopHistory	Identifies the actions and sequence of actions by parties who created, enhanced or transformed the content and/or metadata of the Item.(See Transaction History <hopHistory> below)
Copyright Holder	rightsInfo/copyrightHolder	Party claiming, or responsible for, the intellectual property for the <i>content</i> .
Copyright Notice	rightsInfo/copyrightNotice	The appropriate copyright text claiming the intellectual property for the <i>content</i> . The

		copyright notice may be refined by the Usage Terms.
Item Provider	<code>itemMeta/provider</code>	The party responsible for the <i>management and the release</i> of the NewsML-G2 <i>Item</i> . This corresponds to the publisher of the Item.
Item Generator	<code>itemMeta/generator</code>	The name and version of the system or application that generated the NewsML-G2 <i>Item</i> . Where a role IS NOT specified, the Generator Tool applies to the most recent item generation stage.
Content		
Content Creator	<code>contentMeta/creator</code>	Party that created the <i>content</i> . This MAY also be indicated textually in <by>
Information Sources	<code>contentMeta/infoSource</code>	Person or organisation which originated, modified, enhanced, distributed, aggregated or supplied the content or provided some information used to create or enhance the content Typical example for @role: Isrole:origcont: The original <i>content owner</i> .
Content Contributors	<code>contentMeta/contributor</code>	Person or organization that <i>modified or enhanced</i> the <i>content</i> , preferably the name of a person. Typical contributor types represented by @role: The Editor of the <i>content</i> . An additional Reporter of the <i>content</i> .
Creditline	<code>contentMeta/creditline</code>	Free-text expression of the credit(s) for the <i>content</i> ,
Byline	<code>contentMeta/by</code>	Free-text expression of the person or organisation who created the content– see also Content Creator above

27.4 Transaction History <hopHistory>

Content may traverse many organisations and workflows in its journey to an end user. A news organisation may have third-party providers in addition to in-house contributors and editors. There are a number of reasons why it may be necessary to document the actions involved in the evolution of a news object, for example auditing and payment, or for editorial attribution. Hop History provides a machine-readable way of capturing and expressing metadata describing these actions.

Using Hop History, the receiver of a NewsML-G2 Item should be able to answer questions such as the following:

- ❖ Has the object been transformed?
- ❖ Who transformed the object (and when)?
- ❖ Has the object been enhanced?
- ❖ Who enhanced the object (and when)?
- ❖ Did a specific Party perform a specific Action on a specific Target (the object and/or content and/or metadata)?
- ❖ Has the object been processed (any Action or Target) by a specific Party?
- ❖ What sequence of Party(ies) has processed (any Action or Target) the object?

The object of the actions does not have to be a NewsML-G2 Item; the object of the actions may be the content only, or the content plus the metadata of the Item even if its format was different in a previous hop. Hop History usage is subject to these rules:

- ❖ Hop History does not replace any information that may exist in other formal metadata blocks (itemMeta, contentMeta, partMeta). For example: the Content Creator (a Party in the Hop History



with an associated Action of 'created') MUST be the same as the Party identified as contentMeta/creator and the provider MUST ensure these facts are consistent.

- ❖ It MUST be possible to remove the Hop History without degrading the formal metadata blocks described above.

Each Item can have one <hopHistory> element, which contains one or more <hop> elements. Each <hop> may have a @seq attribute to indicate its place in a sequence of Hops, and may optionally have a @timestamp.

Each Hop element may identify the <party> involved in the Hop, either as a @literal, or @qcode, and an unbounded number of <action> elements that describe the action(s) the Party took. Each <action> element may identify the type of action, using a @qcode, and the @target of the action: whether the action was performed on the full item, the content only, or the metadata.

The Scheme URI of the Hop Action CV (referenced by the @qcode of the action element) is:

<http://cv.iptc.org/newscodes/hopaction/> with a recommended alias of "hopaction". The members of the CV are defined as follows:

Code	Definition
created	A Party has created the target of the action
transformed	A Party has transformed the target of the action. Transforming means that an existing target was changed in data format
enhanced	A Party has enhanced the target of the action

The default action – indicated by the absence of the <action> element – is that the Party forwarded the object without making any changes.

The Scheme URI of the Hop Action Target CV (referenced by @target) is:

<http://cv.iptc.org/newscodes/hopactiontarget/> with a recommended scheme alias of "hatarget". The members of the CV are defined as:

Code	Definition
metadata	Any metadata associated with the content of the object
content	The content of the object, which could be present in different renditions.

The default target – indicated by the absence of the @target attribute, is metadata AND content.

In the Listing below, the provider (Acme Media) expresses a Hop History for the News Item as follows:

- ❖ The original *content* was *created* by Thomson Reuters (note the use of EITHER @qcode OR @literal to identify the Party.)
- ❖ The next Party in the sequence, identified as "comp:bwire", *transformed* the *content* into another format, and *enhanced* the *metadata*.
- ❖ The following Party ("comp:acquiremedia") *created* new *metadata*
- ❖ The final Party in the sequence of Hops ("comp:AP"), *enhanced* the *content* and *metadata*

LISTING 29 Hop History

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsmml:acmenews.com:20111125T1205:HOP-HISTORY-EXAMPLE"
  version="3"
  standard="NewsML-G2"
  standardversion="2.18" xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_25.xml" />
```

```
<catalogRef href="http://www.acmenews.com/synd/catalogs/anmcodes.xml" />
<hopHistory>
  <hop seq="1" timestamp="2010-11-10T16:15:00Z">
    <party qcode="comp:TR">
      <name>Thomson Reuters</name>
    </party>
    <action qcode="hopaction:created" target="hatarget:content" />
  </hop>
  <hop seq="2" timestamp="2011-11-10T16:18:00Z">
    <party qcode="comp:bwire" />
    <!-- transformed content e.g. to a different format -->
    <action qcode="hopaction:transformed" target="hatarget:content" />
    <!-- enhanced existing metadata -->
    <action qcode="hopaction:enhanced" target="hatarget:metadata" />
  </hop>
  <hop seq="3" timestamp="2011-11-12T18:15:00Z">
    <party qcode="comp:aqui remedia" />
    <!-- added metadata -->
    <action qcode="hopaction:created" target="hatarget:metadata" />
  </hop>
  <hop seq="4" timestamp="2011-11-12T19:15:00Z">
    <party qcode="comp:AP" />
    <!-- enhanced content e.g. links to the text of the article -->
    <action qcode="hopaction:enhanced" target="hatarget:content" />
    <!-- enhanced existing metadata -->
    <action qcode="hopaction:enhanced" target="hatarget:metadata" />
  </hop>
</hopHistory>
<rightsInfo>
  <copyrightHolder uri="http://www.thomsonreuters.com" />
</rightsInfo>
<itemMeta>
  <itemClass qcode="ni nat:text" />
  <provider qcode="nprov:reuters" />
  <versionCreated>2011-11-12T20:15:00Z
</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<!-- metadata and content -->
<!-- metadata and content -->
<!-- metadata and content -->
</newsItem>
```

27.5 Hash Value <hash>

The optional <hash> element, when added by a provider, contains a digital “fingerprint” generated from content, that enables end users to verify that the content of a NewsML-G2 Item, or an object referenced by an Item, has not been changed

The <hash> property has two attributes: the @hashtype tells the receiver the algorithm that was used to generate the hash, and @hashscope tells the receiver the parts of the content that were used to create the hash, and that therefore can be verified.

There are two recommended IPTC CVs for the <hash> properties. The first is Hash Type NewsCodes; Scheme URI <http://cv.iptc.org/newscodes/hashtype/> with a recommended scheme alias of “htype”:

Code	Definition
MD5	The MD5 hash function
SHA-1	The SHA-1 hash function
SHA-2	The SHA-2 hash function

The second is the Hash Scope NewsCodes; Scheme URI <http://cv.iptc.org/newscodes/hashscope/> with a recommended scheme alias of “hscope”. Scheme values are:

Code	Definition
content	Content only (Default case, if @hashscope omitted)
provmix	Provider specific mix of content and metadata



In the example below, the scope of the <hash> is a provider-specific mix of metadata and the payload of a News Item. The content would be contained in either an <inline> or <inlineXML> element, and it would be the responsibility of the provider to define (outside NewsML-G2) which metadata properties were in scope for creating the hash. For example they may advise customers that the <headline> and <by> of a NewsItem are included when generating the hash value of the content.

```
<itemMeta>
  ...
  <hash hashtype="htype:md5" hashscope="hscope:provmix">
    hash-code
  </hash>
  ...
</itemMeta>
```

A <hash> element may also be added to <remoteContent> indicating that the hash was generated from the content of the referenced object. In the example below, the target is the thumbnail of an image. The hash scope is omitted and defaults to "content":

```
<remoteContent residref="urn:foobar" rendition="rnd:thumbnail">
  ...
  <hash hashtype="htype:md5">hash-code</hash>
  ...
</remoteContent>
```

When applying a hash value for the content of a NewsML-G2 document, it makes sense to place the <hash> in <itemMeta> only for content payloads of <inlineXML> or <inlineData>. For Remote Content, where there could be multiple renditions of the content, the <hash> should be a child of each <remoteContent> wrapper.



This page intentionally blank



28 Changes to G2-Standards Family

28.1 Introduction

The following is a summary of changes to the G2 Specification since Revision 1 of the Guidelines. Please check the Specification documents available from the IPTC web site (www.newsml-g2.org). The latest changes are in the topmost section, followed by earlier changes in descending order. Details of the changes list below may also be found at <http://dev.iptc.org/G2-Approved-Changes>

28.2 Guidelines release 7 – covers NewsML-G2 2.16 --> NewsML-G2 2.18

Please see 2 [What's New in this release: NewsML-G2 2.18](#)

28.3 Guidelines release 6 – covers NewsML-G2 2.13 --> NewsML-G2 2.15

28.3.1 Add layoutorientation to contentCharacteristics

Expresses editorial advice about the use of a picture in a page layout

28.3.2 Add signal to remoteContent

Enables specific processing instructions for separate renditions of remote content

28.3.3 Add pubconstraint attribute

Applies a constraint to the publication of the value of a metadata property or content reference

28.3.4 Add contenttypevariant attribute

Refines the @contenttype (MIME type) of the referenced resource

28.3.5 Change data type of duration

By changing the data type from integer to string, enables the expression of values such as SMPTE time codes

28.3.6 Add a colourdepth attribute to News Content Characteristics

Expresses the colour depth in bits.

28.3.7 Add a value format attribute to altId

Enables a provider to specify the format of an altId that may be understood outside this NewsML-G2 instance.

28.4 Guidelines release 5 – covered NewsML-G2 2.10 --> NewsML-G2 2.12

28.4.1 Add a link property to rightsInfo

The link may be used by the receiver to access another rights expression resource

28.4.2 Refine AddressType

Opening the cardinality of <area> and <locality> child elements to be unbounded, and add a @role to <line>, <area> and <locality>

28.4.3 Expand geoAreaDetails

Adding three new child elements that express different geometries for defining a Geographical area: <line>, <circle>, <polygon>, each with one or more <position> child elements.



28.4.4 Expressing Ticker Symbols

Adds the ability to express stock prices in a structure; see [Expressing company financial information](#)

28.4.5 Add sameAs to Relationship Group of properties

By adding a repeatable child element <sameAs> to the properties <broadener>, <narrower> and <related>, a list of concept URIs can be provided in addition to the one expressed in the @qcode/@uri of these elements.

Example use: if using <broadener> concept to express the country “parent” of a geographic region, a provider can use <sameAs> to express both two-letter and three-letter country codes in the same wrapper.

28.4.6 Extension of <infoSource>

See [27 Identifying Sources and Workflow Actors](#)

28.4.7 Adding descriptive properties to a package group

By adding <title>, <signal> and <edNote> as child elements of <groupRef>, descriptive metadata and special processing instructions may be provided with the Group References of Package Items. See <http://dev.iptc.org/G2-CR00142-Adding-descriptive-properties-to-a-package-group> for more details.

28.4.8 Add hascontent attribute

See [26.9](#)

28.4.9 New attributes for keyword, subject

See [Section 25.5.2](#)

28.4.10 New <derivedFrom> element

See [Section 25.6](#)

28.4.11 Extend the attributes of an icon

Extending the attributes of an icon in order to enable the selection of the appropriate icon for a specific purpose. See <http://dev.iptc.org/G2-CR00144-extend-the-attributes-of-an-Icon>

28.4.12 Add @uri to properties

See [Section 14.11 A generic way to express concept identifiers as URIs](#)

28.4.13 Ratings of different kinds

See [26.8.1](#)

28.4.14 Adding scheme description to catalog

See <http://dev.iptc.org/G2-CR00149-Adding-scheme-description-to-catalog>

28.4.15 Common set of Power attributes

See [Section 25.5.1](#)

28.4.16 Modify design of XML Schema

See <http://dev.iptc.org/G2-CR00151-modify-design-of-XML-Schema>

28.4.17 Add @title to targetResource properties group

Aligns G2 elements with similar properties provided by HTML5 and Atom. See <http://dev.iptc.org/G2-CR00152-adding-title-attribute-to-targetResource-group>



28.5 Guidelines release 4 – covered NewsML-G2 v2.7 -EventsML-G2 1.6 →NewsML-G2/EventsML-G2 2.9

28.5.1 Unification of NewsML-G2 and EventsML-G2

Please refer to [Chapter 5](#) on [page 23](#).

28.5.2 Hint and Extension Point

Adding properties from the NewsML-G2 (NAR) namespace is a method of providing processing and metadata hints, for example conveying the caption of a remote picture enables this to be displayed without loading the picture itself. However, providing hints in a “flat” list without their parent wrapper element could cause ambiguities, so the inclusion of NewsML-G2 properties at the Extension Point must use the following rules:

- ❖ Any immediate child element from <itemMeta>, <contentMeta> or <concept> may be added directly as a Hint and Extension Point without its parent element;
- ❖ All other elements MUST be wrapped by their parent element(s), excluding the root element.
- ❖ When inserting properties from a target NewsML-G2 resource as processing hints, the properties do NOT have to be extracted directly from the target resource, but they MUST be consistent with the structure of the target resource.

28.5.3 Hop History

Add a machine-readable transaction log an Item. See [27.4](#)

28.5.4 Concept Reference

Enable Event Concepts to be referenced within NewsML-G2 Packages. See [Quick Start - Packages](#) and [EventsML-G2](#)

28.5.5 Extend News Message Header

Enable News Messages to express more semantically-rich properties using QCode values. See [18.2.1](#)

28.5.6 Hash value

Enable the receiver of a NewsML-G2 Item to determine whether the content has been altered. See [27.5](#)

28.5.7 Extend Icon attributes

Add properties of <icon> to enable a basis for choice when multiple icons for content are available. See [10.2.3](#)

28.5.8 Video Frame Rate datatype

Change the datatype of the Video Frame Rate attribute of the News Content Characteristics group from XML Positive Integer to XML Decimal, so that drop-frame frame rates (e.g. 29.97) can be expressed. See [10.2.7.3](#)

28.5.9 Add video scaling attribute

Add an attribute to the News Content Characteristics group to indicate how the original content was scaled to fit the aspect ratio of a rendition. See [10.2.7.5](#)

28.5.10 Colour v B/W image and video

Add an attribute to the News Content Characteristics group to indicate whether the content is colour or black-and-white. See [10.2.7.7](#)



28.5.11 Indicating HD/SD content

Add an attribute to the News Content Characteristics group to indicate whether the content is HD or SD (a non-technical “branded” description of the video content). See [10.2.7.6](#)

28.5.12 Content Warning Best Practice

Add Best Practice for expressing a warning about content and for adding refinements to the warning.

28.5.13 Expressing updates and corrections

Best Practice and structures for expressing that a previous Item has been updated or corrected by the received version

The changes below were covered by Guidelines revision 3 and earlier:

28.6 News Architecture (NAR) v1.7 → v1.8

28.6.1 Redesign planning information: the Planning Item

A new Planning Item is added to the set of G2 items for the purpose of providing information about planned coverage and distributed deliverables independently of the definition of an event.

All G2 Items, including the NewsML-G2 News Items, get an element added under item metadata to reference to Planning Items under which control this item was created and distributed.

This feature is fully documented in [Editorial Planning – the Planning Item](#).

28.6.2 Add a @significance attribute to the <bit> elements of a <bag>

This attribute is assigned to a special use case of a bag with **subject** properties: the bag includes one bit representing an event and one or more other bits representing entities which are related to this event. Only in this case the significance attribute may be used to express the significance of this event to the concept of the bit carrying this attribute.

See [Using @significance with <bag>](#) for a use case and sample code.

28.6.3 Persistent values of @id

A new attribute is added to the Link1Type of property (PCL) that meets the requirement in some circumstances for an @id to be persistent throughout the lifetime of a G2 Item.

For example, the new generic <deliverableOf> element of all G2 items points to a specific <newsCoverage> element of a Planning Item. It is required to make the value persistent over time and versions in order to assure consistency in pointing to the same <newsCoverage> element in different versions of an Item

The change also enables all elements of the Link1Type to point to specific renditions within a News Item, by making the @ids of <inlineXML>, <inlineData> and <remoteContent> persistent in the same way.

28.6.4 Changes to @literal

Implementers commented that the definition on the use of @literal identifiers in some properties was too strict and did not take account of some use cases, specifically:

- ❖ literal values could be defined in a provider’s controlled vocabulary which is defined by other means than G2
- ❖ the absence of @qcode or <bag > should not mandate the use of a @literal.

Therefore some statements about the use of @literal have been redefined and some more precise statements added to avoid ambiguity.

All statements on the literal value in G2 documents should comply with these rules and guidelines:

- ❖ If a literal value is not used with an assert property then it is not required that all instances of that literal value in that item identify the same concept.

- ❖ If a literal value is used with an assert property then all instances of that literal value in that item must identify the same concept.
- ❖ If a <bag> is used with a property then @qcode and @literal attributes must not be used with the property.
- ❖ If a <bag> is not used with a property then the property may have a @qcode attribute or a @literal attribute or **neither**.

Literals may be used as in the following cases:

4. When a code from a vocabulary which is known to the provider and the recipient is used without a reference to the vocabulary. The details of the vocabulary are, in this case, communicated outside of G2. Such a contract could express that a specific vocabulary of literals is used with a specific property.
5. When importing metadata the values of literals may contain codes which have not yet been checked to be from an identified vocabulary.
6. As an identifier for linking with an assert element. In this case the value could be a random one. If a literal value is used with an assert property then all instances of that literal value in that item must identify the same concept.

28.6.5 Deprecate <facet>, extend <related>

The <facet> and <related> properties describe the relationship between two concepts, with <facet> describing an “intrinsic” property of a concept. In practice, it was found that no clear distinction between the use of <facet> and <related> could be made

It was therefore decided to simplify the standard by deprecating <facet> and using only <related>, which has also been extended to express arbitrary values as well as content relationships.

The new features of <related> are documented in [Relationships between Concepts](#).

28.6.6 Extend <contentMeta> for Concept Items and Knowledge Items

This change creates a group of core descriptive metadata properties for G2 items that are available for use with the Planning Item, Concept Item and Knowledge Item.

This makes the Descriptive Metadata Core Group under <contentMeta> consistent across the Planning Item, Concept Item and Knowledge Item. In NAR versions prior to 1.8, the Concept Item has no Descriptive Metadata and the Knowledge Item has a more limited set. The properties are set out below:

Descriptive Metadata Core Group

Title	Name	Cardinality
Language	language	0..unbounded
Keyword	keyword	0..unbounded
Subject	subject	0..unbounded
Slugline	slugline	0..unbounded
Headline	headline	0..unbounded
Description	description	0..unbounded

28.6.7 Add concept details to make them more consistent across concept types

This change aims at applying a consistent design to the details of the different concept types: all should have a date when it was created and one when it ceased to exist.

- ❖ Add a <founded> and a <dissolved> property to <geoAreaDetails>.
- ❖ Add a <ceasedToExist> property to <objectDetails>.
- ❖ Add a <created> and a <ceasedToExist> property to <POIDetails>.

28.7 NewsML-G2 v2.6 → v2.7

No specific changes; inherits all appropriate changes from NAR v1.8



28.8 EventsML-G2 v1.5 → v1.6

No specific changes; inherits all appropriate changes from NAR v1.8

28.9 News Architecture (NAR) v1.6 → v1.7

28.9.1 Move <partMeta> from NewsML-G2 to NAR and extend it

NAR 1.6 moves <partMeta> from being a “NewsML-G2 only” element to inclusion in the overall News Architecture (NAR for G2) framework so that it is available to be used by other G2 Items in addition to the News Item, specifically Package Items and Knowledge Items, including EventsML-G2 Knowledge Items.

The change also extended <partMeta>, adding a @contentRefs attribute, so that it can reference any element in the content section of an item which is identified by @id. For a News Item, these are the child elements of <contentSet>; for the Package Item, the child elements of <groupSet>; and for the Knowledge Item, the child elements of <conceptSet>.

Accompanying this change, the structure of <concept> was also modified to include an optional @id property, enabling it to be referenced from a <partMeta> element of a Knowledge Item.

A use case for this feature is documented in [Handling updates to Knowledge Items using](#)

28.9.2 Extended Rights Information <rightsInfo>

The <rightsInfo> wrapper is extended to enable providers to differentiate between the rights to content and the rights to metadata.

This feature is fully documented, including process models, in [Rights Metadata](#).

28.10 NewsML-G2 v2.5 → v2.6

No specific changes; inherits all appropriate changes from NAR v1.7

28.11 EventsML-G2 v1.4 → v1.5

No specific changes; inherits all appropriate changes from NAR v1.7

28.12 News Architecture (NAR) v1.5 → v1.6

28.12.1 Hierarchy Info element <hierarchyInfo>

The <hierarchyInfo> element was added as a child of <concept> to indicate the position of the concept in a hierarchical taxonomy tree using a sequence of QCodes indicating the ancestor concepts to the left of the target concept. The element is available at CCL and PCL

For example, From the Media Topic NewsCodes (alias=“mtp”) using assumed codes: The concept “adoption” has QCode “mtp:2788”.

Its parent is the concept “family” with the QCode “mtp:2780”

The parent of “family” is the top level concept “society” with the Qcode “mtp:1400”

The resulting Hierarchy Info value is:

`<hierarchyInfo>mtp: 1400 mtp: 2780 mtp: 2788</hierarchyInfo>`

28.12.2 Hint and Extension Point

The G2 design provides for XML Extension Points, allowing elements from any other namespaces, and in some cases also from the NAR namespace, to be added to a G2 element. These Extension Points are now termed “Hint and Extension Points”.

Adding properties from the NAR namespace is a method of providing processing and metadata hints, for example conveying the caption of a remote picture enables this to be displayed without loading the picture itself. In NAR 1.5 a change allows any immediate child element from <itemMeta> or <contentMeta> to be added directly as a Hint and Extension Point without its parent element.



In NAR 1.6, this rule is amended to include the <concept> wrapper. The rule for this feature is now re-stated as follows:

- ❖ Immediate child properties of <itemMeta>, <contentMeta> or <concept> – optionally with their descendants – may be used directly under the Hint and Extension Point.
- ❖ All other properties require the full path excluding only the item's root element.

All other elements MUST be wrapped by their parent element(s), excluding the root element.

28.12.3 Add address details to a Point of Interest (POI)

Up to NAR 1.5 only a position expressed in latitude and longitude values is available to define the location of a POI. In NAR 1.6, a postal address is added to add flexibility to the method of giving details of a location. Note that the address of an organisation given in <contactInfo> may well be different to the actual location of the POI associated with the organisations, e.g. the New York Met location on the map is different to the postal address used for correspondence.

28.12.4 Add <sameAs> to the scheme declarations of a catalog

This feature is fully documented in [14.8 Private versions and extensions of CVs](#)

28.13 NewsML-G2 v2.4 → v2.5

No specific changes; inherits all appropriate changes from NAR v1.6

28.14 EventsML-G2 v1.3 → v1.4

No specific changes; inherits all appropriate changes from NAR v1.6

28.15 News Architecture (NAR) v1.4 → v1.5

The following changes are inherited by NewsML-G2 2.4 and EventsML-G2 1.3

28.15.1 @rendition

The content wrappers <inlineXML>, <inlineData> and <remoteContent> may appear multiple times under <contentSet>, each having a @rendition attribute as processing hint. For example, a picture may have three renditions: “web”, “preview” and “highres”.

To avoid ambiguity, the G2 Specification allows a specific rendition value to be used only once per News Item, i.e. there could not be two “highres” renditions in a content set.

28.15.2 <assert>

The original intention of <assert> was to allow the details of a concept occurring multiple times within a G2 Item to be merged into a single place. However, it was realised that <assert> could also be used to convey rich details of a concept for properties that provided only a limited set of details: name, definition and note.

However, prior to NAR 1.5, the <assert> wrapper could only identify an inline concept using @qcode., whereas a concept can be identified by both @qcode and @literal.

This limitation was removed and <assert> may have EITHER a @qcode or @literal identifier.

28.15.3 Hint and Extension Point

The G2 design provides for XML Extension Points, allowing elements from any other namespaces, and in some cases also from the NAR namespace, to be added to a G2 element. These Extension Points are now termed “Hint and Extension Points”.

Adding properties from the NAR namespace is a method of providing processing and metadata hints, for example conveying the caption of a remote picture enables this to be displayed without loading the picture itself. Prior to NAR 1.5, hints extracted from a target G2 resource could be used freely, i.e. without the need for their parent wrapper element. However, providing hints in a “flat” list could cause ambiguities.



In NAR 1.5 the inclusion of G2 properties at the Extension Point is according to the following rule:

- ❖ Any immediate child element from <itemMeta> or <contentMeta> may be added directly as a Hint and Extension Point without its parent element;
- ❖ All other elements MUST be wrapped by their parent element(s), excluding the root element.

28.15.4 Scheme Code Encoding

A full processing model for Scheme URIs and QCodes was defined. See [14.9](#)

28.15.5 Add @rendition to the icon property

The <icon> property is a child of <link> or <remoteContent> which identifies an image to be used as an iconic identifier for the target resource. If the target resource has multiple renditions, it makes sense to identify which rendition to use for the <icon>

28.15.6 Ranking Multiple Elements

Up to NAR 1.5, the elements that support a @rank attribute are:

- ❖ <link>
- ❖ <broader> | <narrower> | <sameAs>
- ❖ <itemRef>

NAR 1.5 adds the ability to add @rank to the members of the Descriptive Metadata Group, allowing properties such as <language> and <headline> to be ranked by the provider according to an importance that is defined by the provider.

28.15.7 Keyword property

A specific Keyword property was added in NAR 1.5. One reason for the addition was to provide backward compatibility with standards such as IPTC7901, IIM and NITF, which provide a keyword property.

The semantics of keyword are somewhat open: some providers use keywords to denote “key” words that can be used by text-based search engines; some use “keyword” to categorise the content using mnemonics, amongst other examples.

Therefore IPTC suggests the following rules when implementing the Keyword property:

- ❖ Assess if any existing G2 properties align to the use of the metadata. Typical examples are
- ❖ Genres (“Feature”, “Obituary”, “Portrait”, etc.)
- ❖ Media types (“Photo”, “Video”, “Podcast” etc.)
- ❖ Products/services by which the content is distributed
- ❖ If the metadata expresses the subject of the content it could go into the <subject> property with the keyword string itself in a @literal attribute, but it may be better expressed if the keyword string is placed in a <name> child element of the subject with a language tag if required.
- ❖ If migrated to <subject> property, providers should also consider:
- ❖ Adding @type if the nature of the concept expressed by the keyword can be determined
- ❖ Using a QCode if there is a corresponding concept in a controlled vocabulary
- ❖ If none of the above conditions are met, then implementers should default to using the <keyword> property with a @role if possible to define the semantic of the keywords.

The contents of the Keywords field in the example shown below have blurred application: they could properly be regarded as subjects, but the provider intends that they be used as natural-language “key” words that can be used by a text-based search engine to index the content:

```
<keyword role="krole:index">us</keyword>
<keyword role="krole:index">military</keyword>
<keyword role="krole:index">aviation</keyword>
<keyword role="krole:index">crash</keyword>
<keyword role="krole:index">fire</keyword>
```

28.15.8 Multiple Generators

Up to NAR 1.5 only one <generator> per G2 Item is permitted. The use-case is that in some applications where Items are being transformed, a history of <generator> information needs to be preserved, each instance being refined by a @role attribute.

28.15.9 Cardinality of Icon

More than one <icon> property may be given as a child of <contentMeta> or <partMeta> in order to support different renditions, MIME types or formats of the same visual appearance of a target resource icon.

28.16 NewsML-G2 v2.3 → v2.4

Inherits changes from NAR v1.5 plus the following changes specific to NewsML-G2 2.4

28.16.1 New dimension unit indicators for visual content

Some elements holding or referring to news content have the dimension-related attributes Image Height (@height) and Image Width (@width) which are currently defined to be the “number of pixels” of the content dimension. However, some content types require non-pixel units, such as ‘points’ for Graphics; analog video uses different units for Image Width and Image Height.

Therefore in NewsML-G2 2.4 additional attributes have been added to define the Width Unit (@widthunit) and Height Unit (@heightunit). These attributes have QCode values, and the mandatory IPTC CV is

<http://cv.iptc.org/newscodes/dimensionunit/>

The following table shows the default dimension units per visual content type:

Content Type	Height Unit (default)	Width Unit (default)
Picture	pixels	pixels
Graphic: Still / Animated	points	points
Video (Analog)	lines	pixels
Video (Digital)	pixels	pixels

The following example uses the implicit default dimension unit of pixels for a still image:

```
<remoteContent
  resIdref="tag:reuters.com,0000:binary_BTRE4A31LE800-THUMBNAI L"
  rendition="rend:thumbnail"
  contentType="image/jpeg"
  format="fmt:jpegBaseline"
  width="100"
  height="100"
/>
```

The following example uses explicit dimension units:

```
<remoteContent
  resIdref="tag:reuters.com,0000:binary_BTRE37913MM00-THUMBNAI L"
  rendition="rend:thumbnail"
  contentType="image/gif"
  format="fmt:gif87a"
  width="100" widthunit="dimensionunit:points"
  height="100" heightunit="dimensionunit:points"
/>
```

28.17 EventsML-G2 v1.2 → v1.3

Inherits the changes from NAR v1.5. No other changes.

28.18 News Architecture (NAR) 1.3 → 1.4

The following changes are inherited by NewsML-G2 2.3 and EventsML-G2 1.2



28.18.1 Revised Embargo

An embargo can now have an undefined date and time. See [Embargo](#)

28.18.2 New Remote Info element

The <remoteInfo> wrapper is a child of <concept>. It complements the <link> child of <itemMeta> in allowing the creation of links to supplementary resources. Remote Info was added to <concept> so that this information is held within the <concept> structure and therefore retained if the Concept is extracted from the Concept Item and conveyed in a Knowledge Item. See [12.7 Relationships between Concepts](#)

28.19 NewsML-G2 v2.2 → v2.3

Inherits the changes described in [28.2](#), plus the following changes specific to NewsML-G2 2.3

28.19.1 Add <role> to partMeta.

This is in order to indicate the role that part of the content identified by the parent <partMeta> has within the overall content stream. (e.g. "sting". "slate")

28.19.2 Revised Time Delimiter

The <timeDelim> property provides information about the start and end timestamps for parts of streamed content. The @timeunit attribute identifies the units used for the timestamps, defined by the mandatory IPTC Scheme whose URI is <http://cv.iptc.org/newscodes/timeunit/>. In NewsML-G2 2.3, new values were added to the Scheme to cater for additional business requirements that were identified by members.

See [10.6.1](#)

28.19.3 Revised @duration property and new @durationunit

The duration property was defined as the duration in seconds of audio-visual content, but in practice it was found that sub-second precision for measuring time duration was required. The revised definition expresses the duration in the time unit indicated by the new @durationunit.

The duration unit attribute uses the integer value time units of the recommended IPTC Scheme (URI: <http://cv.iptc.org/newscodes/timeunit/>), e.g. seconds, frames, milliseconds, defaulting to seconds if omitted.

See [10.2.7.1](#)

28.20 EventsML-G2 v1.1 → v1.2

Inherits the changes from NAR v1.4. No other changes

28.21 SportsML-G2 v2.0 → v2.1

A number of detailed changes were made to the "plug-in" schemas for individual sports, such as Ice Hockey, Basketball, Tennis and Baseball. Details of these can be found at:

<http://www.iptc.org/std/SportsML/2.1/documentation/sportsml-2.1-changes-additions.html>

A new Tournament Structure was added that will allow implementers to precisely express the Format, Group Stage and Standings of tournaments such as the 2010 FIFA World Cup.

A structure for Series Scores and Results enables the status of a playoff or tournament series to be expressed. Details of the new Tournament Structure are documented at:

<http://www.iptc.org/std/SportsML/2.1/documentation/tournament-structure.html>.



29 Additional Resources

The IPTC plans to refurbish its web site in 2015, so please be aware that the arrangement and location of resources will change after the publication of these Guidelines. However, the web address www.newsml-g2.org will continue to lead you to the landing page about NewsML-G2. The site is divided into logical areas of interest with links to download documentation packages and to web resources.