



IPTC Standards



Guide for Implementers

Including



Public Release

Document Revision 4.0

International Press Telecommunications Council
Copyright © 2011. All Rights Reserved
www.iptc.org



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 © 2011 IPTC, the International Press Telecommunications Council. All Rights Reserved.

This document is published under the Creative Commons Attribution 3.0 license - see the full license agreement at <http://creativecommons.org/licenses/by/3.0/>. 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 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.iptc.org and following the link to News Exchange Formats. 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

Postal address

20 Garrick Street

London WC2E 9BT

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 members who contributed to this documentation project (ordered by surname):

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
GULIJA, Darko	HINA (Croatia News Agency)
HARMAN, Paul	The Press Association
KELLY, Paul	XML Team
LE MEUR, Laurent	Agence France Presse
LINDGREN, Johan	Tidningarnas Telegrambyrå, (Swedish News Agency)
LORENZEN, Jayson	Business Wire
MYLES, Stuart	Associated Press
RATHJE, Kalle	Deutsche Presse Agentur
SCHMIDT-NIA, Robert	Deutsche Presse Agentur
SERGENT, Benoît	European Broadcasting Union
STEIDL, Michael	Managing Director, IPTC
WHESTON, Siobhán	BBC Monitoring
WOLF, Misha	Thomson Reuters

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

About the G2 Standards

The Standards covered by this document are:

- ❖ NewsML-G2, version 2.9
- ❖ EventsML-G2, version 2.9
- ❖ SportsML-G2, version 2.1
- ❖ The News Architecture (NAR) upon which these Standards are based is version 1.10

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

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 “snippets” are shown thus:

```
| standardversion="2.9"
```

Complete XML listings use colours to aid legibility...

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

... and have been validated against the appropriate schema(s).

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>
```

NOT

```
| <inlineXml >
```

Attribute names are always all lowercase For example

```
| standard="NewsML-G2" standardversion="2.9"
```



Indicates especially important notes or cautions to implementers

The term “Item” with capitalised “I” indicates a NewsML-G2 Item (News Item, Planning Item, Package Item etc).

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.

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. These values MUST be present in NewsML-G2 property values based on xs:time or xs:dateTime datatypes, since the exchange of news information is in many cases potentially across different time zones, and timestamps and other time-based properties MUST be unambiguous. The following property values comply with NewsML-G2:

```
| <versionCreated>2010-11-06T12:12:12+01:00</versionCreated>      offset +
| <versionCreated>2010-11-06T12:12:12-01:00</versionCreated>      offset -
| <versionCreated>2010-11-06T12:12:12Z</versionCreated>          indicates UTC
```

The following s NOT compliant

```
| <versionCreated>2010-11-06T12:12:12</versionCreated>          no time offset
```



Table of Contents

Introduction	2
Purpose and Audience	2
Terms of Use.....	2
Contacting the IPTC.....	2
Acknowledgments	3
About the G2 Standards.....	3
Document History.....	3
Conventions used by this document.....	4
Table of Contents	5
Table of Figures.....	14
Table of Code Listings.....	15
1 Executive Summary.....	17
1.1 Why standards matter.....	17
1.2 The purpose of NewsML-G2	17
1.3 Business Drivers.....	17
1.4 Business Requirements.....	18
1.5 Design and Benefits	18
2 About the IPTC.....	19
3 Unification of NewsML-G2 and EventsML-G2.....	19
4 Advice on using the Guide	21
4.1 Note on Code Listings.....	21
5 How News Happens	23
5.1 Introduction.....	23
5.2 Assignment	23
5.3 Information Gathering.....	23
5.4 Verification	24
5.5 Dissemination.....	24
5.6 Archiving.....	24
6 Anatomy of NewsML-G2	27
6.1 Data model.....	27
6.2 Metadata.....	28
6.2.1 NewsML-G2 Properties.....	28
6.2.2 Controlled Values	29
6.2.3 Concepts.....	29
6.2.4 IPTC NewsCodes	29
6.2.5 Qualified Codes (QCodes).....	29
6.3 Conformance Levels.....	31
6.4 Introductory example: a simple News Item.....	32



6.4.1	The top level <newsItem>	33
6.4.2	Item Metadata <itemMeta>	35
6.4.3	Content Metadata <contentMeta>	36
6.4.4	The Content <contentSet>	37
7	Text	39
7.1	Introduction	39
7.2	Example	39
7.3	The top level	39
7.3.1	<newsItem>	39
7.3.2	Catalogs	41
7.3.3	Rights information	41
7.3.4	Summary	41
7.4	Item Metadata	42
7.4.1	<itemClass>	42
7.4.2	<provider>	42
7.4.3	<versionCreated>	42
7.4.4	Optional Management Metadata elements	42
7.4.5	Links	43
7.4.6	Summary	43
7.5	Content Metadata	43
7.6	The Content	45
7.7	Putting it together	45
7.8	Other content options	46
7.8.1	Inline XML	46
7.8.2	Inline data	47
8	Pictures and Graphics	49
8.1	Introduction	49
8.1.1	Use Case	49
8.2	Photo metadata	49
8.2.2	Example Metadata	50
8.3	Photo data	55
8.3.1	Remote Content wrapper	56
8.3.2	News Content Attributes	56
8.3.3	Target Resource Attributes	57
8.3.4	News Content Characteristics	57
8.3.5	Content Hints	58
8.4	The Complete Picture	58
9	Audio and Video	61
9.1	Introduction	61
9.2	Use Case 1: Simple Video for Broadcast	61
9.2.1	Metadata	61



9.2.2	Content.....	65
9.3	Use case 2: Multiple renditions of web video and audio.....	69
10	Packages.....	71
10.1	Introduction.....	71
10.2	Conveying Package Structures.....	73
10.3	Simple Package Structure.....	73
10.3.1	Group Set <groupSet>.....	74
10.3.2	Item Reference <itemRef>.....	74
10.3.3	References to Concepts <conceptRef>.....	76
10.3.4	Additional Properties of <group>.....	77
10.4	Hierarchical Package Structure.....	77
10.5	List Type Package Structure (ordered, bag, alternative).....	79
10.6	Package Processing Considerations.....	81
10.6.1	Other NewsML-G2 Items.....	81
10.6.2	Facilitating the Exchange of Packages.....	82
10.6.3	A “Top Ten” Package Example.....	83
10.6.4	Packaging Other Resources.....	85
10.7	Packages and Links.....	86
10.7.1	The difference in a nutshell.....	86
10.7.2	Purpose and features of the <packageItem>.....	86
10.7.3	When to use <link>.....	86
10.7.4	Link properties.....	86
10.7.5	Link examples.....	88
11	Concepts and Concept Items.....	91
11.1	Introduction.....	91
11.2	What is a Concept?.....	91
11.3	Concept Item.....	91
11.4	Creating Concepts – the <concept> element.....	92
11.4.1	Concept ID <conceptId>.....	92
11.4.2	Concept Name <name>.....	92
11.4.3	Concept Type <type> and Related <related>.....	93
11.4.4	Concept Definition <definition>.....	94
11.4.5	Note <note>.....	95
11.4.6	Remote Info <remoteInfo>.....	95
11.5	Relationships between Concepts.....	95
11.6	Entity Details.....	96
11.6.1	Person Details <personDetails>.....	96
11.6.2	Organisation Details <organisationDetails>.....	98
11.6.3	Geopolitical Area Details <geoAreaDetails>.....	98
11.6.4	Point of Interest <poiDetails>.....	99
11.6.5	Object Details <objectDetails>.....	99



11.7	Code listings.....	100
11.8	Supplementary information about a Concept	102
11.9	Concepts and Controlled Values.....	103
11.9.1	Concept Identifier and Concept URI.....	103
11.9.2	From Concept URI to QCodes	104
11.9.3	Resolution of QCodes.....	104
11.10	Concepts in Practice	105
12	Knowledge Items	107
12.1	Introduction	107
12.2	Structure and Properties.....	108
12.2.1	The <knowledgeItem> element.....	109
12.2.2	Management Metadata.....	109
12.2.3	Content Metadata.....	110
12.2.4	Concept Set	110
12.3	Use Case: a Controlled Vocabulary for <accessStatus>	111
13	Controlled Vocabularies and QCodes	115
13.1	Introduction	115
13.2	Business Case	115
13.3	How QCodes work	115
13.3.1	Scheme Alias to Scheme URI	116
13.3.2	Concept URI.....	116
13.4	QCodes and Taxonomies	118
13.5	Managing Controlled Vocabularies as NewsML-G2 Schemes.....	119
13.5.1	Knowledge Items	119
13.5.2	Creating a new Scheme	119
13.5.3	Creating a new Knowledge Item for distributing a CV	119
13.5.4	Maintaining a CV	120
13.6	Processing Controlled Vocabularies.....	121
13.6.1	Resolving Scheme Aliases.....	121
13.6.2	Resolving Concept URIs	122
13.7	Private versions and extensions of CVs.....	125
13.7.1	Solution for Case 1	125
13.7.2	Solution for Case 2	126
13.8	QCodes and non-URI characters.....	128
13.8.1	Non-ASCII characters	128
13.8.2	White Space in Codes.....	128
13.9	Syntactic Processing of QCodes	129
13.9.1	Creating QCodes from Scheme Aliases and Codes	129
13.9.2	Processing Received Codes	129
13.10	QCode and Literal Identifiers.....	130
13.10.1	The difference in a nutshell.....	130



13.10.2	What @literal is; what it is NOT	130
13.10.3	Use of @literal in a NewsML-G2 Item.....	130
13.10.4	Changes to @literal	131
14	EventsML-G2	133
14.1	A standard for exchanging news event information.....	133
14.1.1	Business Advantages.....	133
14.1.2	Structure and compatibility	134
14.2	Event Information – What, Where, When and Who	136
14.2.1	What is the Event?.....	136
14.2.2	Where does it take place?.....	136
14.2.3	When does it take place?.....	137
14.2.4	Who is involved?	137
14.3	Event Coverage.....	137
14.4	Event Properties.....	137
14.4.1	Event Description	138
14.4.2	Event Relationships.....	139
14.4.3	Event Details Group.....	140
14.5	Event Use Cases	150
14.5.1	Event in a Concept Item.....	151
14.5.2	Multiple Event Concepts in a Knowledge Item.....	154
14.5.3	Packaging Events using <conceptRef>	158
14.5.4	Events in a News Item	159
14.5.5	Adding event concept details to a News Item.....	161
15	Editorial Planning – the Planning Item.....	165
15.1	Item Metadata <itemMeta>	165
15.2	Content Metadata <contentMeta>	166
15.3	<newsCoverageSet>	166
15.4	<newsCoverage>	166
15.5	<planning>.....	167
15.5.1	<planning> at Power Conformance Level.....	167
15.6	The <delivery> component.....	171
15.7	Delivered Items - <deliveredItemRef>	171
15.7.1	Delivered Item Reference: Core Conformance	171
15.7.2	Delivered Item Reference: Power Conformance.....	173
16	SportsML-G2	175
16.1	Introduction	175
16.2	Business Advantages	175
16.3	Structure	175
16.3.1	XML Schema and Catalogs	176
16.3.2	Item Metadata.....	177
16.3.3	Content Metadata.....	177



16.3.4	InlineXML and the Sports Content component.....	178
16.3.5	Sports Event wrapper in more detail.....	179
16.3.6	SportsML-G2 Packages.....	184
17	Exchanging News: News Messages.....	191
17.1	Introduction	191
17.2	Structure	191
17.2.1	News Message Header <header>	192
17.2.2	News Message payload <itemSet>	195
18	Migrating IPTC 7901 to NewsML-G2.....	197
18.1	Introduction	197
18.1.1	Message Header	197
18.1.2	Message Text	198
18.1.3	Post-Text Information.....	198
19	News Industry Text Format (NITF).....	201
19.1	Introduction	201
19.2	Overview of NITF and NewsML-G2 equivalents.....	202
19.2.1	Root element <nitf>	202
19.2.2	Document metadata <head>	202
19.3	Content metadata <body.head>	204
19.4	Body Content <body.content>.....	205
19.5	End of Body <body.end>	205
20	Mapping Embedded Photo Metadata to NewsML-G2	207
20.1	Introduction	207
20.2	IPTC Metadata for XMP	208
20.3	Synchronizing XMP and legacy metadata.....	208
20.4	Picture services using IIM/XMP.....	208
20.5	Rationale for moving to G2.....	208
20.6	IIM resources	209
20.7	Approach	209
20.8	IIM to NewsML-G2 Field Mapping Reference.....	209
20.9	IIM to NewsML-G2 Mapping Example.....	210
20.9.1	Object Name / Title.....	211
20.9.2	Urgency.....	212
20.9.3	Category, Supplementary Category.....	212
20.9.4	Keywords.....	212
20.9.5	Special Instruction.....	213
20.9.6	Date Created, Time Created.....	213
20.9.7	By-line, Credit, Source.....	213
20.9.8	City, Province/State and Country.....	214
20.9.9	Original Transmission Reference.....	215
20.9.10	Headline	216



20.9.11	Copyright Notice	216
20.9.12	Caption/Abstract.....	216
20.9.13	Writer/Editor.....	216
20.10	Reconciling NewsML-G2 with embedded metadata.....	217
21	Rights Metadata.....	219
21.1	Rights Info (Core Conformance Level).....	219
21.2	Business Reasons for Extended Rights Information at PCL	219
21.3	The extended <rightsInfo> model	220
21.3.1	Rights Info Scope NewsCodes	220
21.3.2	Rights Aspect NewsCodes.....	220
21.4	Use cases for the extended <rightsInfo> model	221
21.4.1	Separating rights for “content” and “metadata”.....	221
21.4.2	Separating rights for part of the metadata.....	221
21.4.3	Separating rights for aspects of part of the metadata.....	222
21.4.4	Expressing rights to different parts of content.....	222
21.4.5	Expressing rights to aspects of content.....	223
21.5	Processing model for extended Rights Info.....	224
21.5.1	Adding Rights Information expressions: the news provider’s view.....	224
21.5.2	Processing Diagram: Rights Holder view	227
21.5.3	Processing Diagram: Content User view.....	229
22	Advanced Metadata Techniques	231
22.1	Introduction	231
22.2	The Assert wrapper	231
22.2.1	Business cases	231
22.2.2	Using Assert to map from IIM or XMP.....	234
22.3	Inline Reference.....	235
22.3.1	Quantify Attributes Group	235
22.3.2	Linking text content to an Inline Reference.....	236
22.3.3	Using <inlineRef> and <assert> together	236
23	Generic Processes and Conventions.....	239
23.1	Introduction	239
23.2	Processing Rules for NewsML-G2 Items	239
23.3	Publishing Status	239
23.4	Embargo.....	241
23.5	Geographical Location	243
23.6	Processing Updates and Corrections.....	245
23.6.1	Example Use Case using <link> to previous version and @severity.....	245
23.7	Content Warning	247
24	Verification and Security	249
24.1	Introduction	249
24.2	Transaction History <hopHistory>	249



24.3	Hash Value <hash>	251
25	Changes to G2-Standards Family.....	253
25.1	Introduction	253
25.2	NewsML-G2 v2.7-EventsML-G2 1.6 →NewsML-G2/EventsML-G2 2.9.....	253
25.2.1	Unification of NewsML-G2 and EventsML-G2	253
25.2.2	Hint and Extension Point.....	253
25.2.3	Hop History	253
25.2.4	Concept Reference.....	253
25.2.5	Extend News Message Header.....	253
25.2.6	Hash value.....	253
25.2.7	Extend Icon attributes.....	253
25.2.8	Video Frame Rate datatype.....	253
25.2.9	Add video scaling attribute.....	253
25.2.10	Colour v B/W image and video	254
25.2.11	Indicating HD/SD content.....	254
25.2.12	Content Warning Best Practice	254
25.2.13	Expressing updates and corrections	254
25.3	News Architecture (NAR) v1.7 → v1.8	254
25.3.1	Redesign planning information: the Planning Item	254
25.3.2	Add a @significance attribute to the <bit> elements of a <bag>.....	254
25.3.3	Persistent values of @id.....	254
25.3.4	Changes to @literal.....	254
25.3.5	Deprecate <facet>, extend <related>	255
25.3.6	Extend <contentMeta> for Concept Items and Knowledge Items	255
25.3.7	Add concept details to make them more consistent across concept types	255
25.4	NewsML-G2 v2.6 → v2.7.....	256
25.5	EventsML-G2 v1.5 → v1.6.....	256
25.6	News Architecture (NAR) v1.6 → v1.7	256
25.6.1	Move <partMeta> from NewsML-G2 to NAR and extend it	256
25.6.2	Extended Rights Information <rightsInfo>	256
25.7	NewsML-G2 v2.5 → v2.6.....	256
25.8	EventsML-G2 v1.4 → v1.5.....	256
25.9	News Architecture (NAR) v1.5 → v1.6	256
25.9.1	Hierarchy Info element <hierarchyInfo>	256
25.9.2	Hint and Extension Point.....	256
25.9.3	Add address details to a Point of Interest (POI).....	257
25.9.4	Add <sameAs> to the scheme declarations of a catalog.....	257
25.10	NewsML-G2 v2.4 → v2.5.....	257
25.11	EventsML-G2 v1.3 → v1.4.....	257
25.12	News Architecture (NAR) v1.4 → v1.5	257
25.12.1	@rendition.....	257



25.12.2	<assert>	257
25.12.3	Hint and Extension Point	257
25.12.4	Scheme Code Encoding	258
25.12.5	Add @rendition to the icon property	258
25.12.6	Ranking Multiple Elements.....	258
25.12.7	Keyword property.....	258
25.12.8	Multiple Generators.....	259
25.12.9	Cardinality of Icon	259
25.13	NewsML-G2 v2.3 → v2.4.....	259
25.13.1	New dimension unit indicators for visual content	259
25.14	EventsML-G2 v1.2 → v1.3.....	259
25.15	News Architecture (NAR) 1.3 → 1.4	259
25.15.1	Revised Embargo	260
25.15.2	New Remote Info element.....	260
25.16	NewsML-G2 v2.2 → v2.3.....	260
25.16.1	Add <role> to partMeta.....	260
25.16.2	Revised Time Delimiter	260
25.16.3	Revised @duration property and new @durationunit.....	260
25.17	EventsML-G2 v1.1 → v1.2.....	260
25.18	SportsML-G2 v2.0 → v2.1.....	260
26	Additional Resources.....	261



Table of Figures

Figure 1: The heart of NewsML-G2 is the News Architecture (NAR); a framework which allows companies to support many different information needs – for example news, events, sports statistics – using common components	18
Figure 2: How the NAR builds into NewsML-G2 instances	27
Figure 3: Browser view of the remote Catalog.....	30
Figure 4: NewsML-G2 News Item derived from the News Architecture	32
Figure 5: One of the IPTC Custom metadata panels	49
Figure 6: Each <remoteContent> wrapper references a separate rendition of the resource.....	56
Figure 7: A “Top Ten” package on the Web Source: Thomson Reuters.....	71
Figure 8; Structural model of a NewsML-G2 Package Item resembles the News Item.....	72
Figure 9: Simple Package using Item References <itemRef>	74
Figure 10: Simple Package Relationship.....	75
Figure 11: Code outline of Hierarchical Package with two Groups.....	78
Figure 12: Relationship diagram of Hierarchical Package.....	78
Figure 13: Using <groupRef> to create an unordered “bag” of groups of Items	80
Figure 14: Groups G1 and G2 are both children of the Root Group.....	80
Figure 15: Diagrams of Package Profiles. The numbers in brackets indicate the required items.....	82
Figure 16: Concept Items share the same template as News Items and Package Items.....	92
Figure 17: Entering the Concept URI in a browser returns human-readable information.....	104
Figure 18: Information Flow for Concepts and Knowledge Items	107
Figure 19: Structure of a Knowledge Item mirrors that of other NewsML-G2 Items	109
Figure 20: Human-readable browser page of a Concept URI.....	117
Figure 21: Persistent Event: An event conveyed as a Concept using <eventDetails>.....	135
Figure 22: Transient Event(s): Multiple event information carried as <events> in a News Item.....	136
Figure 23: Hierarchy of Events created using Event Relationships.....	139
Figure 24: An event represented in a News Item or a Concept Item uses a common structure	151
Figure 25: Event Flow using Package Items and Knowledge Items.....	159
Figure 26: Planning Item: the newsCoverageSet has one or more newsCoverage components	166
Figure 27: Structure of a SportsML-G2 document matches NewsML-G2	176
Figure 28: Diagram of Package Group structure.....	187
Figure 29: News Messages can convey any number of any kind of G2 Item.....	191
Figure 30: Syndication workflow using Destination and Channel	194
Figure 31: One of the IPTC Core metadata panels from an image opened in Adobe Photoshop.	207
Figure 32: Opening File Info panel of the image (<i>photo and metadata</i> ©Thomson Reuters).....	210
Figure 33: How to apply rightsInfo elements referencing only a fraction of a NewsML-G2 Item	226
Figure 34: Processing diagram to identify the parts to which a rightsInfo refers.....	227
Figure 35: Processing diagram showing how to identify the rights associated with a specific element....	229
Figure 36: State Transition Diagram for <pubStatus>	240
Figure 37: Processing <embargoed>.....	242



Table of Code Listings

LISTING 1: Example of a NEWSML-G2 Document	33
LISTING 2: A NEWSML-G2 Document	45
LISTING 3: Conveying a Picture in NewsML-G2	58
LISTING 4: Simple Video in NewsML-G2.....	66
LISTING 5: Multiple renditions of audio/video.....	69
LISTING 6: Simple Group Set at CCL.....	75
LISTING 7: Simple Group Set at PCL	76
LISTING 8: Simple Group Set including <conceptRef>	77
LISTING 9: Hierarchical Package.....	78
LISTING 10: List Type Package.....	81
LISTING 11: The complete Top Ten style News Package.....	83
LISTING 12: Simple Group Set using @href at CCL.....	85
LISTING 13: “Abstract” Concept.....	100
LISTING 14: “Named Entity” Concept.....	101
LISTING 15: Knowledge Item for Access Codes.....	112
LISTING 16: Event sent as a Concept Item.....	152
LISTING 17: Update to the Previously-Sent Event	154
LISTING 18: Two Related Events conveyed in a Knowledge Item.....	156
LISTING 19 A Set of Events carried in a News Item	160
LISTING 20 Planning Item at CCL.....	167
LISTING 21 Planning Item at PCL	171
LISTING 22 Planning Item with delivery at CCL.....	172
LISTING 23 Planning Item with delivery at PCL	173
LISTING 24: A complete sample SportsML-G2 document.....	182
LISTING 25: Sports story in NewsML-G2/NITF.....	184
LISTING 26: SportsML-G2 Package	188
LISTING 27: News Message conveying a complete News Package.....	195
LISTING 28: An NITF marked-up article conveyed in <inlineXML>	205
LISTING 29: Embedded photo metadata fields mapped to NewsML-G2	216
LISTING 30: Illustrating Located, Subject and Dateline	243
LISTING 31: Hop History	250



This page intentionally blank



1 Executive Summary

1.1 Why 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.

1.2 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.

1.3 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.

■ *"The fastest way to move into the rapidly growing digital economy is to adopt standards, which will ... enable businesses to maximise their investments and obtain industry-leading performance at lower cost and with greater choice."*

Craig Barrett, CEO, Intel

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

SportsML-G2 is a markup 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)

1.4 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
- ❖ 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.

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

1.5 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.

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 NewsM-G2 complements the work of companies working on search and navigation technologies to realise the vision of the Semantic Web.

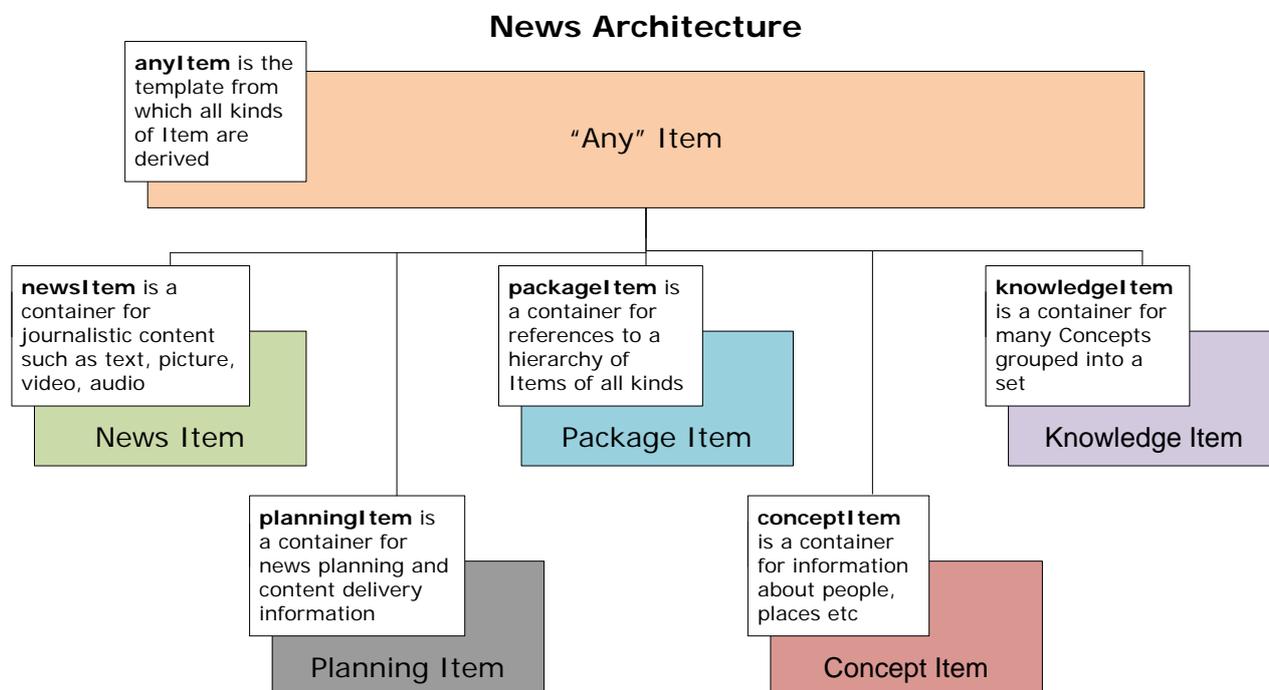


Figure 1: The heart of NewsML-G2 is the News Architecture (NAR); a framework which allows companies to support many different information needs – for example news, events, sports statistics – using common components



2 About the IPTC

The IPTC's members are technologists and decision-makers from the world's main news agencies and leading media players. They are experts 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 are widely used by the news agencies and their newspaper and broadcaster customers, as is the IIM standard for pictures. All media organisations benefit from these standards; they have been a key enabler in 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.

3 Unification of NewsML-G2 and EventsML-G2

The G2 Standards NewsML-G2 and EventsML-G2 have always had much in common; they added only a few standard-specific features to the shared G2 News Architecture. With the introduction of the Planning Item in NewsML-G2 v2.7 and EventsML-G2 v1.6, it became even more difficult to make a distinction between the two separate standards; the Planning Item adopted the News Coverage wrapper element, which in turn was deprecated from the EventsML-G2 <event> structure. It is also possible to have a Planning Item containing references only to NewsML-G2 Items.

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 and there is no management of it as a separate specification. As a result the variants of the concept structure – with and without Event Details (eventDetails) – have been merged into a single structure that includes eventDetails, thus ending the need for NewsML-G2 and EventsML-G2 to have separate Concept and Knowledge Items.

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, and replace the multiplicity of Schema/Schema References that were previously used. (Individual component Schemas still exist and are available via links on the IPTC web site)



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



4 Advice on using the Guide

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 some to hand, such as the G2 Specification, which may contain some further detailed information about features discussed.

Additional Resources details the location of these companion resources. There is also a chapter on **Changes to G2-Standards Family** which summarizes the changes and new features of G2 since Revision 1 of the G2 Implementation Guidelines.

All implementers should be familiar with the contents of the Chapters entitled **How News Happens** and **Anatomy of NewsML-G2**. It may also be helpful to read the Chapter on **Text** since this introduces some of the core structures used throughout the G2 Family.

This background information will enable implementers of **EventsML-G2** and **SportsML-G2** to treat the Chapters dedicated to these topics as “standalone”.

Best practice advice common to all implementations are discussed in Chapters on **Rights Metadata**, **Advanced Metadata Techniques** and **Generic Processes and Conventions**

There are chapters dedicated to using NewsML-G2 to convey **Text, Pictures and Graphics** and **Audio and Video**. Every effort has been made to provide links in cases where features being used are described in more depth elsewhere in the document.

An important feature of NewsML-G2 is the ability to convey “packages” of all types of media in managed relationships; this is detailed in a Chapter on **Packages**.

The role that G2 Standards play in News Planning is covered in the Chapter on **Editorial Planning – the Planning Item**, which convey information about planned news coverage and about items of content that will be, or have been delivered.

Implementers who are migrating from IPTC 7901 or NITF will find specific information in **Migrating IPTC 7901 to NewsML-G2** and **News Industry Text Format (NITF)**. A chapter on

Mapping Embedded Photo Metadata to NewsML-G2 contains advice for converting IIM and embedded application metadata, such as Photoshop XMP, to NewsML-G2

Although optional, it is likely that implementers will be interested in conveying and exchanging knowledge about news and entities found in news. These are details in Chapters on **Concepts and Concept Items** and **Knowledge Items**.

The management of the exchange of news content is detailed in a Chapter on **Exchanging News: News Messages**.

4.1 Note on Code Listings

All of the Code Listings have been validated against the appropriate XML Schema. The schema location is assumed to be on the same local path as the XML file. Implementers who wish to use the code must install the appropriate schema(s) on a local file path of their choosing and amend the code accordingly.

The relevant schemas may be downloaded from the IPTC Web site www.iptc.org. Follow the links to News Exchange Formats → NewsML-G2 (or SportsML-G2) → Specification to obtain a link to download a ZIP package of schema files, documentation and examples.

Implementers are recommended to use the “All-Core” Schema for Core Conformance, or the “All-Power” Schema for the Power Conformance Level. These cover all Item types in both NewsML-G2 and EventsML-G2, and replace the multiplicity of Schema/Schema References that were previously used. (Individual component Schemas still exist and are available via links on the IPTC web site)

For SportsML-G2, a link to ZIP package containing schema, documentation and examples can be found at the SportsML pages of the IPTC web site.



Implementers are requested NOT to validate documents directly against the IPTC's hosted schemas. If validation is required, implementers MUST validate against locally-stored copies of the appropriate IPTC Schemas.



5 How News Happens

5.1 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

5.2 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.

It is these business processes that [EventsML-G2](#) and [Editorial Planning – the Planning Item](#) address.

5.3 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.

5.4 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.

5.5 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, ("canceled" 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 are available, many organisations currently 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.

5.6 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.



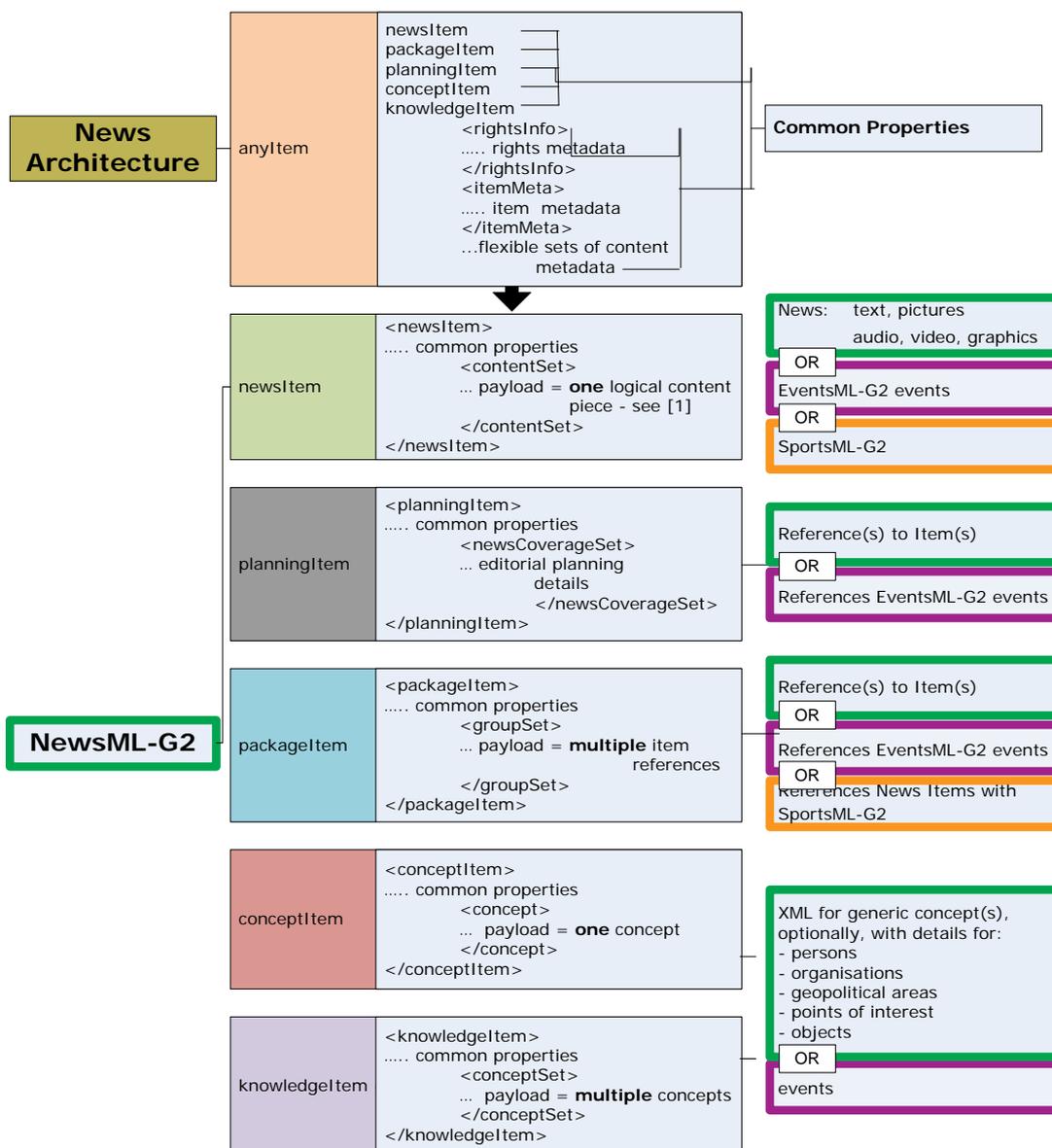
This page intentionally blank

6 Anatomy of NewsML-G2

6.1 Data model

NewsML-G2 documents instantiate a specific class of *Item*, depending on required application of the standard, as shown below:

- ❖ the newsItem is a generic container for any type of news content;
- ❖ the planningItem conveys information relating to the planning and fulfilment of news.
- ❖ the packageItem contains grouped references to many NewsML-G2 Items;
- ❖ conceptItem and knowledgeItem express concepts and collections of concepts



[1] Alternative renditions of the same content are permitted, e.g. a picture as JPEG, GIF and TIF in a single newsItem

Figure 2: How the NAR builds into NewsML-G2 instances

SportsML-G2 and EventsML-G2 markup is always conveyed in a NewsML-G2 document structure, The News Architecture itself is independent of any specific technical implementation. This Guide describes the implementation of the IPTC-provided XML Schema, for the purpose of exchanging information, but the model could be implemented using object-oriented software or in a database..

The objective of the NAR is to model the structure, management and associations of news content in a manner that conforms to the principal notions of the Semantic Web.

6.2 Metadata

One of the rules of NewsML-G2, and a clear advantage for implementers, is the separation of content and metadata. In an IPTC7901 text message, there is a limited amount of metadata which is separated from the content by simple means using spaces and line-feeds. This can be a limitation in today's multi-channel news environment.

For example, on a web page the printed words of an article are the content. This content has certain properties, or **metadata**, including the URL of the page, the date of publication, the section (navigation) and links with other content that may be regarded as **administrative metadata**.

The content itself may have properties such as a headline by-line and abstract, or summary. These are often seen as part of the **structure** of the content, but may also be regarded as part of the **descriptive metadata**.

In order to tell a consumer how to locate the article, we need to use as much of this metadata as possible. Other useful properties may be keywords, or categories, which enable the article to be grouped with other similar articles.

In a digital environment, this rudimentary metadata is no longer sufficient: the volume of content available means we must have more metadata in order that users can find what they need in a mass of competing information. The more metadata a publisher can provide, the more likely it is that users will view their content, and not that of their competitors. NewsML-G2 Properties offer powerful support for metadata that can be used to create links to the content.

■ *"In a link-and-search economy, content gains value only through these recommendations; an article without links has no readers and thus no value."*
Jeff Jarvis, columnist, The Guardian

6.2.1 NewsML-G2 Properties

A core design feature of NewsML-G2 is to make the standard relatively straightforward to maintain and expand. To this end, properties are kept as generic as possible, consistent with making the standards easy to implement and understand.

For example, rather than have separate properties to indicate, people, organisations and places, NewsML-G2 uses the generic Subject property, and qualifies the nature of the subject using an attribute. This means that any type of entity may be classified, without having to define a new type of property and issue a revised schema.

Each property is based on one of a number of re-usable templates

6.2.1.1 String, Block and Label Type

These types of property hold natural language text that is intended to be human-readable.

6.2.1.2 Qualified Properties

These are properties that may only hold a value from a *Controlled Vocabulary* (see **Controlled Values**)

6.2.1.3 Flexible Properties

May hold either a controlled or uncontrolled (literal) value.

6.2.1.4 Other Property Types

There are a number of specialised property types for expressing date-time, numbers, IRIs and links

6.2.1.5 Property Groups

Some properties are notionally organised into groups. There are two kinds of groups, element groups (e.g. Concept Definition Group) and attribute groups (e.g. News Content Characteristics Group).



Note that these groups have no formal significance; they are merely constructs within the XML Schema to facilitate maintenance and as an aid to implementers.

6.2.2 Controlled Values

Values for metadata can be controlled, or uncontrolled – and it is often desirable for metadata values to be controlled, that is, restricted to a value or range of values.

One obvious reason for doing this is to convey clear and unambiguous information about content. If a provider needs to inform a customer that the content is a photograph, what term should be used: photograph, photo, picture, pic? They might be understood by a human reader, but ad hoc terms may not be processed reliably by software. In the same way that a drop-down list in a GUI restricts user choice to a known range of values, we need a Controlled Vocabulary (CV) to ensure a standard term is used that all receivers and applications can reliably process, and possibly in multiple languages.

CV is a generic term that encompasses our notions of a coding scheme, taxonomy, drop-down menu, thesaurus, dictionary etc. NewsML-G2 provides this functionality in a lightweight and portable manner.

6.2.3 Concepts

When creating or consuming news, we almost always need to express something more about the content, such as some categorisation. Increasingly, we try to identify objects in the news, such as the names of people involved, and provide links to other information about these entities.

In the Semantic Web, all of these things are modelled as relationships, for example an article *belongs to* a certain category of news.

Concepts are the generic term used in NewsML-G2 to denote real-world entities, such as people, organisations and places, and also abstract notions such as subject categories, facial expressions. Concepts are a model for managing this information and making it available via CVs, enabling a single piece of news content to be linked to a network of information resources.

There is a detailed discussion on managing Concepts in [Concepts and Concept Items](#).

6.2.4 IPTC NewsCodes

The IPTC maintains sets of CVs that are collectively branded *NewsCodes*. These represent concepts that describe and categorise news objects in a consistent manner. By standardising on NewsCodes, providers can ensure a common understanding of news content and a greater degree of inter-operability between content from different providers.

For example, news publishers need to tell users whether the content of a news message is text, picture or some other type of object. A piece of free text could be used, but this would become unworkable: if an application received content described as “φωτογραφία”² it may not be able to process the content correctly. Providers use the NewsCode taxonomy that classifies the “nature” of news items in a machine-readable form that can be reliably translated to human-readable information. An example of this concept expressed in a NewsML-G2 message would be:

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

This example also illustrates the use of *QCodes* to identify Concepts.

6.2.5 Qualified Codes (QCodes)

QCodes consist of two parts, separated by a colon (:). In our example, the first part of the QCode “ninat” is an alias that can be used to identify the IPTC NewsCode vocabulary concerned with the nature of content (ninat= *News Item Nature*). We refer to these aliases throughout NewsML-G2 documentation as *Scheme Aliases*. The second part of the QCode “picture” is a reference into the newsItem nature vocabulary.

² “photograph”



Scheme aliases are resolved by looking in *Catalog* information carried at the root level of a NewsML-G2 document. Catalog information can be conveyed inline use the <catalog> element, but the more common approach is to reference online Catalogs, using <catalogRef>, thus:

```
<catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
```

Resolving the URL carried by the above catalogRef in a web browser, shows the following:

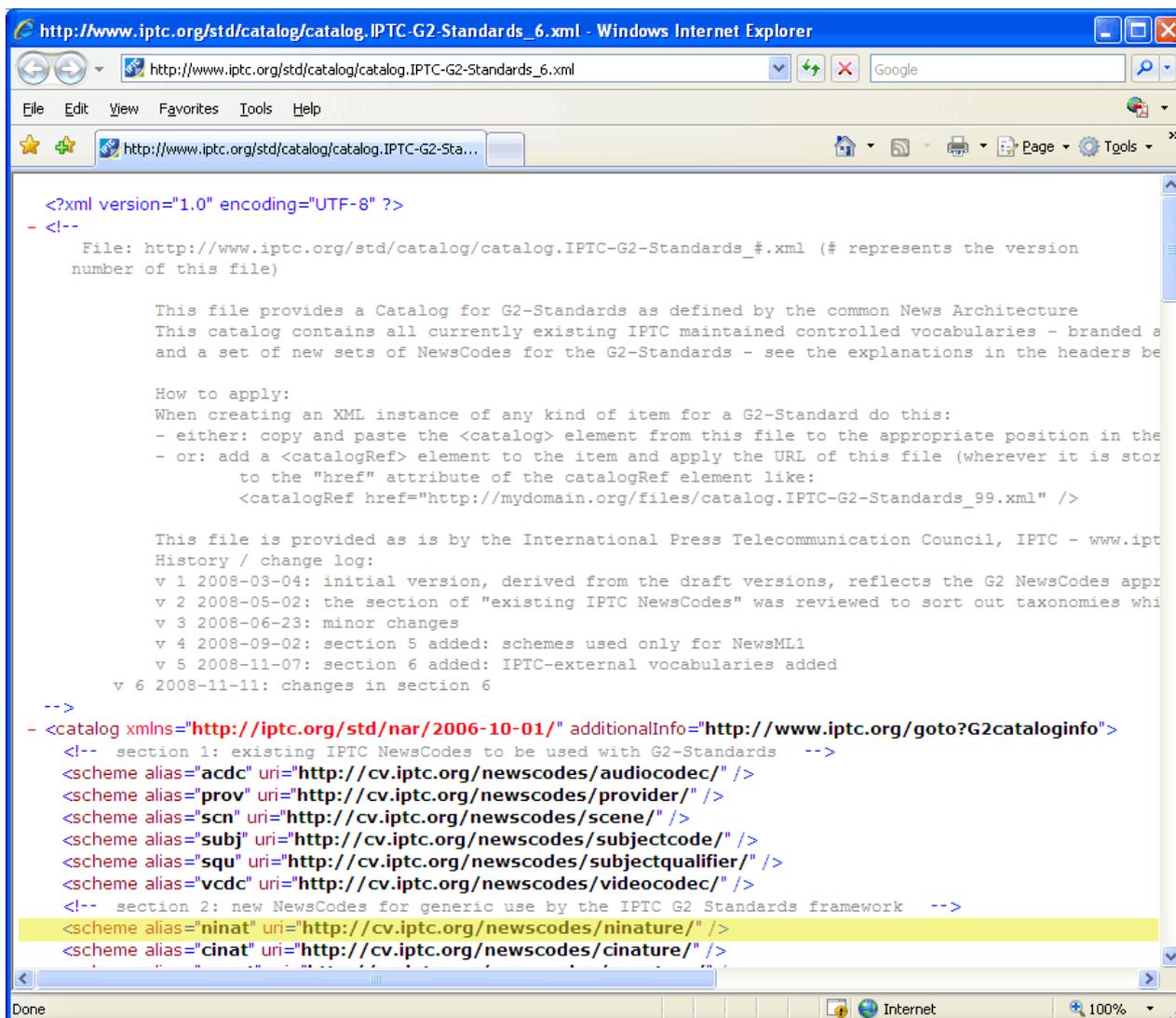


Figure 3: Browser view of the remote Catalog



Highlighted in yellow is the mapping from the “ninat” alias for the NewsCode scheme to the scheme URI. If we look at this location in a web browser, we see in natural-language description of the terms (in English) in the controlled vocabulary³ are as follows:

Concept ID	Definition (en-GB)
ninat:animated	Animated graphic content in a News / Package Item
ninat:audio	Audio Content in a News / Package Item
ninat:composite	Composite content in a News / Package Item
ninat:graphic	Still (unanimated) graphic content in a News/PackageItem
ninat:picture	Picture content in a News/PackageItem
ninat:text	Text content in a News/PackageItem
ninat:video	Video content in a News/PackageItem

Using this technique, property definitions in other languages may be made available, for example (note: not complete):

Concept ID	Definition (gr)
ninat:animated	διακινούνται τέχνης περιεχόμενο
ninat:audio	Το περιεχόμενο ήχου

NewsML-G2 documents should be reasonably easy for humans to interpret, as well as facilitating automatic processing. Scheme aliases and scheme values can therefore be made meaningful, or at least recognisable, to humans. Technically, there is no reason to do so. A QCode of `<xyz: spaghetti>` may look meaningless, but if it resolves to something that software can process, that is all that the standard requires.

For a more in-depth discussion of QCode processing, please read [Controlled Vocabularies and QCodes](#).

6.2.5.1 QCodes and QCode types

QCodes are used throughout NewsML-G2 documents as a way of expressing controlled values. Apart from the `@qcode` used by many properties, other attributes may have a QCode data type. For example:

```
<subject type="cpnat: abstract" qcode="subj: 01000000" />
```

Both attributes are QCode types; the second one is actually called “qcode”.

6.2.5.2 QCode lists

Sometimes, an implementer needs to be able to express a relationship to more than one Concept within the same property. Multiple QCodes are allowed in the QCode List type of property. Each of the QCode values is separated by a space:

```
<contactInfo role="crol: home crol: private">
```

6.3 Conformance Levels

The NAR defines two conformance levels:

- ❖ Core Conformance Level (CCL) is the default and is focussed on simplicity and inter-operability
- ❖ Power Conformance Level (PCL) provides a set of extensions to the Core model that gives additional features, but inter-operability is likely to be lower, since not all recipients will implement the Power level.

However, implementers are free to choose which Power features they need; they are not forced to use PCL for all properties, only those for which the power features are required.

³ Note that this particular piece of information does not convey the format or encoding of the content, which is described separately.

Thus, an application implemented at CCL SHOULD be able to process a PCL document, but simply ignore any Power features.

If a NewsML-G2 document implements PCL features, it MUST declare the conformance level in the root level of the Item. If the conformance property is missing, CCL is assumed.:

| conformance="power"

6.4 Introductory example: a simple News Item

The News Item conveys a single piece of journalistic content covering an event considered newsworthy by the provider, in the form of the written word, picture, graphic, or some other digital media.

The structure of a NewsML-G2 "Item" – in this case the News Item – consists of four parts, as shown below in **Figure 4**:

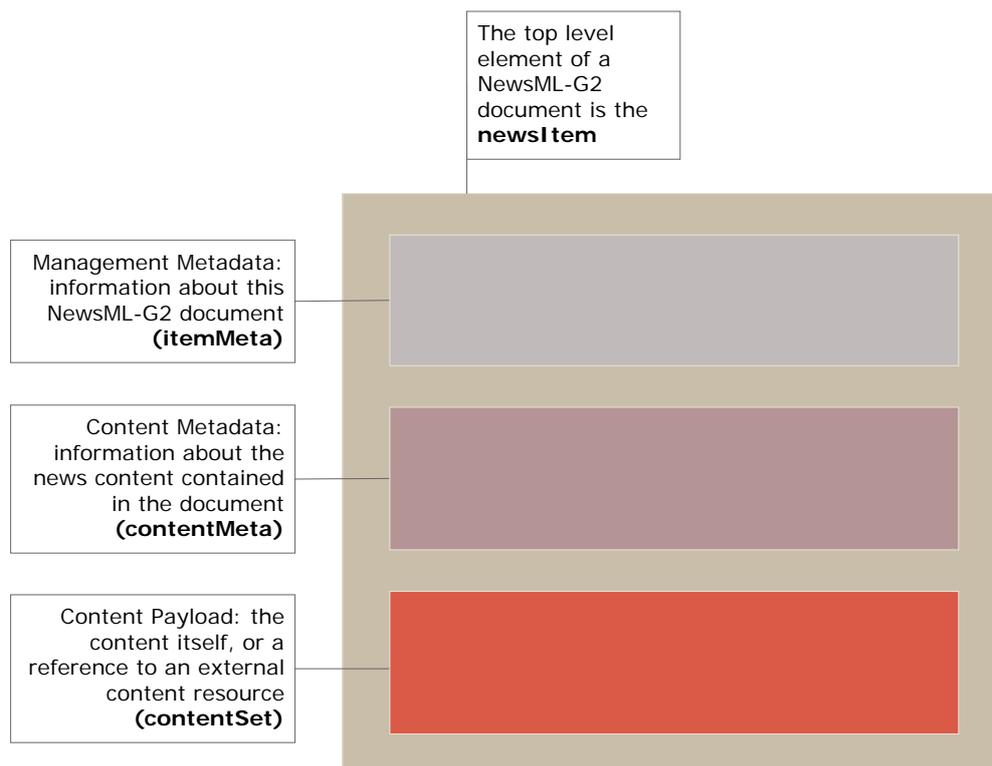


Figure 4: NewsML-G2 News Item derived from the News Architecture

- ❖ The top-level element <newsItem> contains a globally unique id (GUID) for the document, XML namespace(s), catalog reference(s) copyright and licensing information.
- ❖ Management metadata <itemMeta> conveys expressions about the NewsML-G2 documents and properties needed for the appropriate handling of the content, such as the itemClass (what type of content is conveyed by the Item), and whether it is usable.
- ❖ Content Metadata <contentMeta> contains expressions about the content: *administrative* metadata such as the information source, timestamps; *descriptive* metadata such as the category of news etc.
- ❖ Content can be any media type contained in the <contentSet> wrapper. Character content, e.g. story text, may be included in-line in a chosen format, such as XHTML and NITF (News Industry Text Format, see www.nitf.org). Small binary objects may also be included inline. Binary content – such as a picture, video or graphic – would normally be included by reference to an external resource.

LISTING 1 below illustrates a simple (but not minimum) NewsML-G2 document, with the structural elements emphasised in colour according to the structural diagram shown in **Figure 4**, followed by a description of the each of the lines of XML.



LISTING 1 Example of a NEWSML-G2 Document

```
<?xml version="1.0" encoding="UTF-8" ?>
<newsItem guid="urn:newsml:iptc.org:20101106:tutorial-text-xhtml" version="1"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="core"
  xml:lang="en-GB" >
  <catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml"
  />
  <itemMeta>
    <itemClass qcode="inat:text" />
    <provider literal="IPTC" />
    <versionCreated>2010-11-06T12:12:12+01:00</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <contentCreated>2010-11-05</contentCreated>
    <located literal="Roma">
      <name>Rome</name>
    </located>
    <creator literal="MSteidl">
      <name>Michael Steidl</name>
    </creator>
    <language tag="en-GB" />
    <subject type="cpnat:abstract" qcode="subj:01000000" />
    <subject type="cpnat:abstract" qcode="subj:01026000" />
    <subject type="cpnat:abstract" qcode="subj:01026002">
      <name>art, culture and entertainment &gt; news &gt; media</name>
    </subject>
    <subject type="cpnat:organization" literal="IPTC">
    <slugline>IPTC-NewsML-G2</slugline>
    <headline>Latest version of NewsML-G2 is approved</headline>
  </contentMeta>
  <contentSet>
    <inlineXML contenttype="application/xhtml+xml">
      <html xmlns="http://www.w3.org/1999/xhtml">
        <head>
          <title>IPTC-NewsML-G2</title>
        </head>
        <body>
          <h1>NewsML-G2 version approved by the IPTC in Rome</h1>
          <p>
            <span class="dateline">Rome, November 5 (IPTC) -</span>
            The International Press Telecommunications Council Standards Committee
            today approved a new developer version of its News Exchange Format
            NewsML-G2.
          </p>
        </body>
      </html>
    </inlineXML>
  </contentSet>
</newsItem>
```

[<return to Code Listings index](#)

6.4.1 The top level <newsItem>

The first line is standard XML

```
<?xml version="1.0" encoding="UTF-8" ?>
```

Next, the top-level element <newsItem> contains a mandatory globally unique identifier (GUID) for the NewsML-G2 document, and, optionally, a version number:

```
guid="urn:newsml:iptc.org:20101106:tutorial-text-xhtml" version="1"
```

The GUID must be globally unique for all time. The syntax shown in this example uses the URN (Uniform Resource Name) registered by the IPTC. (See [GUID](#)) but providers are free to use their own scheme. One popular choice is the Tag URI scheme (see [TAG URI home page](#))



The IPTC namespace and, additionally, the W3C namespace for XML Schema:

```
xml ns="http://iptc.org/std/nar/2006-10-01/"
xml ns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

6.4.1.1 Schema Reference:

```
xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/XSD/NewsML-G2_2.9-spec-All-Core.xsd"
```

This locates the IPTC’s schema for NewsML-G2 version 2.9 on the local file system and associates it with the IPTC namespace.



Implementers need to consider whether there is a real need to validate NewsML-G2 documents in a production environment – assuming that documents are created and processed by software. To do so also creates an external dependency with possible performance and service implications.

For cases where validation is required, implementers MUST validate against locally-stored copies of the appropriate IPTC Schemas.

6.4.1.2 Standard and version:

```
standard="NewsML-G2" standardversion="2.9"
```

The Standard and Standard Version values MUST be aligned with the NewsML-G2 Schema being used. This Implementation Guide is aligned with version 2.9 of NewsML-G2 and as of this version forward, the value of @standard should always be “NewsML-G2”.

For prior versions, before the unification of EventsML-G2 under a single set of NewsML-G2 schemas, the permitted values for @standard are “NewsML-G2” or “EventsML-G2” Documents should therefore be processed according to the table below.

@standard	@standardversion	Note
NewsML-G2	2.0	
EventsML-G2	1.0	
NewsML-G2	2.1	
NewsML-G2	2.2	
EventsML-G2	1.1	
NewsML-G2	2.3	
EventsML-G2	1.2	
NewsML-G2	2.4	
EventsML-G2	1.3	
NewsML-G2	2.5	
EventsML-G2	1.4	
NewsML-G2	2.6	
EventsML-G2	1.5	
NewsML-G2	2.7	
EventsML-G2	1.6	
NewsML-G2	2.8	
EventsML-G2	1.7	
NewsML-G2	2.9	For ALL documents in the G2 Family

The values of these attributes enable a processor to determine which XML Schema is being used by the document, in preference to a direct schema reference which, as noted above, would have performance and service implications in a high-volume production environment. From NewsML-G2 v2.9 onwards there is a single XML Schema each for “core” and “power” implementations of NewsML-G2 and EventsML-G2 (See [Unification of NewsML-G2 and EventsML-G2](#))



6.4.1.3 Conformance

If the conformance level of the Item is PCL, this MUST also be stated, otherwise CCL is assumed. In the example, the conformance level is “core” (note that this is the default, and could be omitted)

```
| conformance="core"
```

6.4.1.4 Catalog reference

```
| <catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml"/>
```

This identifies the IPTC NewsCodes catalog, and will enable a processor to de-reference QCodes as required. Any other catalogs used by the document must be declared in this section.

The IPTC recommends that a processor caches any remote catalogs indefinitely, so that the catalog provider’s servers are not overloaded with requests, and to avoid processing dependencies in production.

Optionally, we have declared that the default language used in the text of the document is UK English (but note that this only applies to the XML of the document - the content itself may be in other languages and uses a separate property in content metadata)

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

The <newsItem> may also contain a set of rights and licensing information in a <rightsInfo> element. We will show this in later examples and in more detail in a section on [Rights Metadata](#).

6.4.2 Item Metadata <itemMeta>

Item Metadata is information about the <newsItem> as a whole and how it should be processed, There are three mandatory elements: <itemClass>, <provider> and <versionCreated>.

In the first line of our example, a QCode using the IPTC’s NewsCode scheme for the News Item Nature tells the recipient that the content of the document is text:

```
| <itemClass qcode="ni nat: text" />
```

(Learn more about NewsCodes and QCodes in [Controlled Vocabularies and QCodes](#))

The next mandatory element identifies the provider of the Item

```
| <provider literal="IPTC" />
```

The <provider> element uses the Flexible Property type template, which means it takes either a QCode or a “literal” value.

The final mandatory element records the date and time (including offset from UTC) that the newsItem (note: not the content) was created as per ISO 8601: YYYY-MM-DDThh:mm:ss±hh:mm (note that NewsML-G2 requires the time offset to be included, or “Z” to indicate UTC).

```
| <versionCreated>  
| 2010-11-06T12:12:12+01:00  
| </versionCreated>
```

A <pubStatus> property is required, but may be omitted if the status is “usable”. If not present, the item and its contents are assumed to be “usable”. The property has been included in this example:

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

Publication status is likely be used by most public providers, especially news agencies, for whom 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” and this case is set to “usable”. Other values permitted by the scheme are:

- ❖ canceled. (Note the U.S. spelling). This means that the content of the newsItem (or referenced by the newsItem) must not be used, ever.
- ❖ withheld means the content must not be used until further notice.



A typical use-case for cancelling or withholding a News Item is when a provider has already sent content, but needs to cancel or hold it for legal reasons. 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 [Publishing Status](#).

6.4.3 Content Metadata <contentMeta>

This is the information directly associated with the content of the Item and is notionally divided into two groups:

- ❖ Administrative Metadata – information such as the source of the information, the urgency
- ❖ Descriptive Metadata – information about the subject matter of the Item, such as category, a headline.

There are no mandatory elements, but in our example we have shown some which are likely to be used by many providers.

The <contentCreated> property contains a date stamp in YYYY-MM-DD format and optionally the time with time zone. The property is a TruncatedDateTime data type, which allows the date value to 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).

```
<contentCreated>2010-11-05</contentCreated>
```

Next we have a <located> element which carries information about the place that the content was created (note: note necessarily where the event itself took place). Instead of using a controlled vocabulary, we are using a literal ID, such as.⁴

```
<located literal="Roma" />  
  <name>Rome</name>  
</located>
```

In later examples we show how geographical information can be conveyed more precisely.

The creator of the content is identified by a literal value (in this example, but a controlled value may be used):

```
<creator literal="MStei dl" />  
  <name>Mi chael Stei dl</name>  
</creator>
```

The language of the content is specified using the IETF's BCP47 (Best Current Practice). (see [IETF BCP47 Page](#) for further information).

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

Next we have some subject codes to tell the receiver how the provider has categorised the content, in this case according to the IPTC's Subject NewsCodes

```
<subject type="cpnat:abstract" qcode="subj:01000000" />  
<subject type="cpnat:abstract" qcode="subj:01026000" />  
<subject type="cpnat:abstract" qcode="subj:01026002">  
  <name>art, culture and entertainment &gt; news &gt; medi a</name>  
</subject>
```

The subject element as used here has two attributes, type and qcode, both of which are QCodes. The first “cpnat:abstract” references an IPTC NewsCodes scheme which indicates what kind of concept is described by the “subj:01000000” QCode. In this case this code is an abstract concept, i.e. not a real-world entity such as “person” or “organisation”.

This can be an important distinction. For example, an application could identify people associated with a News Item, that have been coded in the metadata, by finding all concepts of type “person” referred to by the document, and automatically produce links to further resources.

⁴ Strictly, this value is an identifier, not a human-readable label. The NewsML-G2 Specification allows a @literal to be used as a label, if usable as such, in the absence of other information, but providers do not have to make @literal identifiers meaningful to the human reader. It is recommended that a <name> property should be used in these circumstances.



The following line illustrates:

```
<subject type="cpnat:organisation" literal="IPTC" />
```

Here the “concept nature” (cpnat:) is “organisation” and the subject is a literal value “IPTC” rather than a QCode.

The slugline is used by many content providers to informally identify content:

```
<slugline>IPTC- NewsML G2</slugline>
```

And finally in the <contentMeta> section we have a headline element.

```
<headline>NewsML- G2 v2.7 is approved</headline>
```

Note that we may also have a headline in the content itself – some providers like to deliver the headline in the article, others do not. The <contentMeta> headline may be also used independently of any content, and for example when the content is binary yet still requires some textual identification.

6.4.4 The Content <contentSet>

In this simple example we show the inclusion of text as inline XML:

```
<inlineXML contentType="application/xhtml+xml">
```

In succeeding chapters we will look at the how any type of content can be carried or referenced by a NewsML-G2 News Item, with uses cases for text, pictures, audio and video, sports statistics, and event planning.



This page intentionally blank



7 Text

7.1 Introduction

In the previous chapter, we looked at a simple NewsML-G2 document from the point of view of a recipient, and discussed the document structure, briefly looking at the elements and attributes used by the document. Now we show some of the essential building blocks of NewsML-G2 in more detail, again using a piece of text content as an example. In subsequent chapters, we cover the specific use cases of pictures, video and audio.

7.2 Example

Below we have 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-Robinson
Created on	2008-11-21 15:21:06
Source	ANM
Author	mjameson
Latest edit	2008-11-21 16:22:45
Latest editor	moiras
Category Codes	TRS, FIN
People	Hank Robinson
Organisations	US Treasury
Headline	Robinson: Must preserve bailout funds
Byline	By Meredith Jameson
Location / Date	Washington 21/11/2008
Body Text	Et, sent luptat luptat, commy nim zzriureet vendreetue modo dolenis ex euisis 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 nulput 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 all of the information needed to create a valid NewsML-G2 document, except for some boilerplate applicable to Acme News and Media.

7.3 The top level

7.3.1 <newsItem>

The top level (root) element <newsItem> MUST contain the following attributes, and in this recommended order:

7.3.1.1 Standard

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

| standard=“NewsML- G2”

7.3.1.2 Standard Version

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

| standardversi on=“2. 9”



7.3.1.3 GUID

Each NewsML-G2 document MUST provide an identifier that is guaranteed to be globally unique for all time and independent of location, (The standard permits documents to be physically located in many places but each document may have only one ID). The form of the identifier MUST be an IRI (Internationalized URI). One way to guarantee uniqueness is to create an ID using some algorithm that combines the provider name and the date-time that the document was first generated, combined with some other information that may be needed to disambiguate the document.

The IPTC has registered a URN for the purpose of creating GUIDs for NewsML-G2 Items using a specification in RFC 3085bis. The syntax for a GUID using this scheme⁵ is:

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

The ProviderId must be in the form of a valid internet domain name owned by the provider: in our case this is acme-news.com. The DateId must be in the form CCYYMMDD, and the NewsItemId must be unique for all items published by the Provider on this Date. For this example we will use the story's Slugline (but note that this is unlikely to be granular enough in a real-world implementation).

```
| guid="urn: newsml : acme- news. com: 20081125: US- FINANCE- ROBINSON"
```

Any ID that can be guaranteed to be globally unique is valid for NewsML-G2. Some implementers use a Tag URI ([more information here](#)), for example:

```
| guid="tag: afp. com, 2008: TX- PAR: 20080529: JYC80"
```

Other schemes used include UUID (Universally Unique Identifier, a URN namespace) (see [IETF RFC 4122 Page](#)) and DOI (Digital Object Identifier).



Where a provider has used some algorithm based on date and time (such as RFC 3085) to create a GUID, recipients MUST NOT "reverse engineer" the information to create a time-stamp. This may have unintended consequences and result in errors. There are a number of date-time properties in NewsML-G2 which accurately convey a provider's intentions.

7.3.1.4 Further Attributes

Conformance. All NewsML-G2 documents have two conformance levels, Core and Power. If no Conformance attribute is present, Core Conformance Level (CCL) is assumed. The Power Conformance Level (PCL) gives implementers more features, but potentially at the cost of loss of interoperability.

```
| conformance="core"
```

Version. Each NewsML-G2 document has a version, which MUST start at 1. If @version is not present, it MUST be assumed to be 1. For each update to a document that has the same GUID, the version number MUST be incremented, not necessarily by 1.

```
| version="1"
```

Language. We will specify U.S. English as the default language for the text properties in the document using BCP47.

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

Dir. Specifies the directionality of the textual content, for example a Hebrew text would be

```
| dir="rtl"  
<!-- right to left -->
```

Namespace and schema information should also be added as appropriate. (see [6.4.1.1](#))

⁵RFC 3085 adds a suffix to the GUID :[RevisionId][Update]. This is not used in NewsML-G2 and the need to use it is removed in RFC 3085bis. The version information is carried by the `version` attribute of `<newsitem>`



7.3.2 Catalogs

On the editing screen, we can see a number of fields which are metadata that the report's author and editors have ensured will go with this story:

Category codes: TRS, FIN. These are typically used by applications to route content to appropriate channels or services. Using controlled vocabularies and QCodes we can make them more accessible to human readers.

People: Henry Robinson. We have faked a link in the editing system which would point to some resource that can provide more information about Henry Robinson – say a picture and biographical information.

Organisations: US Treasury. Again we have faked links to other resources.

If the codes used are QCodes, providers MUST place Catalog information at the top level of the document so that receivers' processing application can resolve them. A Catalog is a set of pointers to taxonomies of subjects, people, organisations, places etc identified by a URL.

Some mandatory NewsML-G2 properties MUST use values from an IPTC NewsCode; these MUST be referenced using EITHER the <catalogRef> element to a remote catalog:

```
<catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
```

OR one can place Catalog information directly in the NewsML-G2 document, using the <catalog> element, which contains the <scheme> child property with @alias and @URI identifying the resource, thus:

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

Most providers will have more than one scheme, and referring to all them in a series of <catalog> statements may make the document more verbose and potentially unreliable (For example if a Catalog statement is inadvertently omitted) Therefore a single <catalogRef> property with an @href to the remote Catalog file containing all of the <catalog> statements may be used. As the document we are creating contains metadata which is proprietary to ANM, we need to reference the remote (please note: fictional) ANM Catalog:

```
<catalogRef href="http://www.acmenews.com/syndication/catalogs/anmcodes.xml" />
```

For more on Catalogs and Schemes, please see [Knowledge Items](#).

7.3.3 Rights information

An optional <rightsInfo> wrapper can hold copyright statements and usage terms. We will place a copyright notice in this example document and take a more detailed look at <rightsInfo> in [Pictures and Graphics](#) and in [Rights Metadata](#).

7.3.4 Summary

The top level <newsItem> element of the example document in full is:

```
<?xml version="1.0" encoding="UTF-8" ?>
<newsItem guid="urn:newsml:acmenews.com:20101225T1205:US-FINANCE-ROBINSON"
  version="3" xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd"
  standard="NewsML-G2"
  standardversion="2.9"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.acmenews.com/synd/catalogs/anmcodes.xml" />
  <rightsInfo>
    <copyrightHolder literal="Acme News and Media LLC" />
  </rightsInfo>
```



7.4 Item Metadata

The <itemMeta> section has three mandatory elements, which must be present in the following order:

7.4.1 <itemClass>

Providers MUST use the IPTC NewsCode denoting the type of information asset that is conveyed by the NewsML-G2 document. In this case, the content payload is text. As seen in the previous example, we express this in the <itemClass> element:

```
<itemClass qcode="ni nat: text" />
```

7.4.2 <provider>

This may take a QCode or a literal value attribute. If using a QCode, the IPTC maintains the Provider NewsCodes, a controlled vocabulary of providers registered with the IPTC. As Acme News & Media are fictional, we will use a literal value:

```
<provider literal="ANM" />
```

7.4.3 <versionCreated>

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

```
<versionCreated>2011-12-25T11:25:32-05:00</versionCreated>
```

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

7.4.4 Optional Management Metadata elements

Element	Element Name	Datatype		Comment
Date Item First Created	firstCreated	DateTime		
Embargo	embargoed	Flexible		See 23.4 Embargo
Publish Status	pubStatus	QCode	Canceled Usable Withheld	MUST use the IPTC Publishing Status NewsCode scheme
Role in Workflow	role	QCode	Flash Urgent Lead Alert Bulletin	Examples are from the IPTC Editorial Role NewsCode scheme, but provider-specific codes are permissible
File Name	filename	String		The recommended file name for this Item
Editorial Service	service	QCode		Provider-specific values for the service(s) or feed(s) on which the item may be carried. Optional child element <name>
Item Title	title	Label		A natural language title
Editorial Note	edNote	Block		A natural language journalistic note addressed to users.

We will use <pubStatus> and <service> in the example.

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



For the Editorial Service, we will assume that ANM has a controlled vocabulary of codes for its services, under the scheme alias “svc” and that the code for the service is “USN”, for “U.S. News”. This is represented in NewsML-G2 as:

```
<service qcode="svc:USN">  
  <name>U. S. News</name>  
</service>
```

The <service> property does not necessarily convey an exhaustive list of all of the services that the item is carried on. It may express a provider’s intention to carry the item on specific services. However, if for example the item is re-distributed automatically by software, there is no requirement for <service> to be updated. To do so may be impractical, or cause unnecessary activity by creating new versions of the item.

7.4.5 Links

NewsML-G2 allows us to express links to supporting or additional resources, and also to identify the location, type and size of the linked resource(s). The relationship of the linked resource to the Item can be described, and the IPTC maintains a set of NewsCodes for this purpose.

For example, this Item may be a translation of another article. An @href to the original article could be provided, with @rel expressing the relationship “Translated From” An example is given in [9.2.1.9](#)

Links can also be useful for providing a trail to previous versions of the Item’s content. In some editorial systems, all articles get a new ID, so a correction to a previously-sent article would not retain the previous version’s ID and incremented version number, but a completely new ID. In these circumstances, a provider could provide the ID of the previous version in a Link, with a @rel expressing the semantic “previous version”. (Discussed further in [Processing Updates and Corrections](#))

7.4.6 Summary

Our complete <itemMeta> section is:

```
<itemMeta>  
  <itemClass qcode="innat:text" />  
  <provider literal="ANM" />  
  <versionCreated>2011-12-25T11:25:32-05:00</versionCreated>  
  <pubStatus qcode="stat:usable" />  
</itemMeta>
```

7.5 Content Metadata

The <contentMeta> section contains information about the content that is carried, or referenced, by the NewsML-G2 document. Conceptually, there are two kinds of property of <contentMeta>: Administrative and Descriptive.

7.5.1.1 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, who created it or helped to create it, where the news coverage took place, and to whom the content should be directed.

We have three set of administrative properties to convey in our sample story:

- ❖ Timestamps
- ❖ Story location
- ❖ Writer

The two timestamps are <contentCreated>, corresponding to a “Created on” field of the story, and <contentModified>, corresponding to a “latest edit” field. These are expressed in NewsML-G2 as Truncated DateTime data type. (See [Content Metadata <contentMeta>](#)) Both are optional, and the only rule is that if both are present, the Modified datetime MUST be greater than (i.e. after) the Created datetime.

```
<contentCreated>2011-12-25</contentCreated>
```



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

```
<located literal="WDC" />  
  <name>Washington DC</name>  
</located>
```

Note that this is not necessarily the same place as the occurrence of the event being reported. A story about a place in the UK may have been written in the London office, in which case the place identified by <located> would be "London", as should be reflected in the natural-language <dateline> property (see also [Geographical Location](#)). Using QCodes, location can be conveyed more precisely, and in terms that may be more readily processed by software.

We can express the Writer of the article using the <creator> element:

```
<creator literal="MJameson" />  
  <name>Meredith Jameson</name>  
</creator>
```

7.5.1.2 Descriptive Metadata

These properties set the context of the news being conveyed in relation to other news items by describing and classifying it, using genre and subject codes.

It also enables us to isolate metadata that was traditionally carried within the content itself such as the headline and by-line, and treat this as metadata. This is an immediate practical benefit, because by doing so we are no longer forced to open or retrieve the content in order to process and use it.

None of the Descriptive Metadata properties is mandatory, but the example story has some which we need to convey:

- ❖ Subject Codes
- ❖ Headline
- ❖ Slugline

Subject properties will use QCodes that are owned and maintained by the provider ANM. Thus:

```
<subject qcode="ANMCat: TRS"
```

tells us that "ANMCat" is the alias for ANM's Category Code scheme, and "TRS" is the reference within the scheme to the "Treasury" category of news. We can also add the natural-language name of the Category using the <name> tag:

```
<subject qcode="ANMCat: TRS">  
  <name>Treasury News</name>  
</subject>  
<subject qcode="ANMCat: FIN">  
  <name>Financial News</name>  
</subject>
```

For Organisations and People, we will use ANM's schemes for identifying these resources. We will also use the IPTC NewsCode to label each of these schemes according to the type of "thing" – or concept – being described, in this case "organisation" and "person". This extra level of classification will help consumers to execute queries such as "find further information about all of the people associated with this story".

```
<subject type="cpnat: organisation" qcode="ANMorg: 5768933">  
  <name>U. S. Treasury Department</name>  
</subject>  
<subject type="cpnat: person" qcode="ANMpers: 9999999">  
  <name>Hank Robi nson, Treasury Secretary</name>  
</subject>
```

"cpnat" is the alias for the IPTC scheme that contains the "nature of the concept". "ANMpers" and ANMorg" are the scheme aliases for ANM's controlled vocabularies of people and organisations, respectively.

The <slugline> property uses the "Slugline" field of the story:

```
<slugline> US- Fi nance- Robi nson</slugline>
```

In a similar fashion, the <headline> property will use the "Headline" field:



```
<headline>Robinson: Must preserve bailout funds</headline>
```

7.6 The Content

The content of the NewsML-G2 document is enclosed by the `<contentSet>` wrapper. In **LISTING 1** we showed a minimal NewsML-G2 document that carried XHTML content. In this example, we will show the content carried as NITF (News Industry Text Format). This is an XML standard, so will be contained by an `<inlineXML>` element, and we will use `@contenttype` to denote the XML vocabulary used, which must be expressed as an IANA MIME type.

```
<contentSet>
  <inlineXML contenttype="application/nitf+xml">
    <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/">
      <body>
        <body.head>
          <headline>
            <hl1> Robinson: Must preserve bailout funds</hl1>
          </headline>
          <byline>
            <byttl>By Meredith Jameson</byttl>
          </byline>
        </body.head>
        <body.content>
          <p>Et, sent luptat luptat, conmy nim zzriureet vendreetue modo
            dolenis ex euisis 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 feugiamconmy nostrud min ulla
            autpat velisl duisismodip ero dipit nit utpatum sandrer cipisim
            nit lortis augiat nulla faccum at am quam velenis nulput la
            augerostrud magna commolore eliquatie exerate facilis
            modiamconsed dion henisse quipit at. Ut la feu facilla feu
            faccumsan ecte modoloreet ad ex el utat.
          </p>
          <p>Ugiating ea feugait utat, venim velent nim quis nulluptat num
            Volorem inci enim dolobor eetuer sendre ercin utpatio dolorpercing
            Et accum nullan voluptat wisis alit dolessim zrrilla conmy nonulpu
            tpatinis exer sequatueros adit verit am nonse exerili quismodion
            esto cons dolutpat, si.
          </p>
        </body.content>
      </body>
    </nitf>
  </inlineXML>
</contentSet>
```

In later chapters we will show how we can wrap alternative renderings of the same content within a `<contentSet>`, using further properties to distinguish between them.

7.7 Putting it together

The complete listing is shown in **LISTING 2** below.

LISTING 2 A NEWSML-G2 Document

```
<?xml version="1.0" encoding="UTF-8" ?>
<newsItem guid="urn:newsml:acmenews.com:20101225T1205:US-FINANCE-ROBINSON"
  version="3" xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd"
  standard="NewsML-G2" standardversion="2.9" xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.acmenews.com/synd/catalogs/anmcodes.xml" />
  <rightsInfo>
    <copyrightHolder literal="Acme News and Media LLC" />
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ninat:text" />
    <provider literal="ANM" />
    <versionCreated>2011-12-25T11:25:32-05:00</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <contentCreated>2011-12-25</contentCreated>
```



```

<located literal="WDC" />
  <name>Washington DC</name>
</located>
<creator literal="MJameson" />
  <name>Meredith Jameson</name>
</creator>
<language tag="en-US" />
<subject qcode="ANMCat: TRS">
  <name>Treasury News</name>
</subject>
<subject qcode="ANMCat: FIN">
  <name>Financial News</name>
</subject>
<subject type="cpnat: organization" qcode="org: 5768933">
  <name>U.S. Treasury Department</name>
</subject>
<subject type="cpnat: person" qcode="pers: 9999999">
  <name>Hank Robinson, Treasury Secretary</name>
</subject>
<slugline>US- Finance- Robinson</slugline>
<headline>Robinson: Must preserve bailout funds</headline>
</contentMeta>
<contentSet>
  <inlineXML contentType="application/nitf+xml">
    <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/">
      <body>
        <body.head>
          <headline>
            <hl1> Robinson: Must preserve bailout funds</hl1>
          </headline>
          <byline>
            <byttl>By Meredith Jameson</byttl>
          </byline>
        </body.head>
        <body.content>
          <p>Et, sent luptat luptat, commy nim zzriureet vendreetue modo
            dolenis ex euisis 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 ulla
            autpat velisl duisismodip ero dipit nit utpatum sandrer cipisim
            nit lortis augiat nulla faccum at am, quam velenis nulput la
            augerostrud magna commolore eliquatie exerate facilis
            modiamconsed di on hennisse quipit at. Ut la feu facilla feu
            faccumsan ecte modoloreet ad ex el utat.
          </p>
          <p>Ugiating ea feugait utat, venim velent nim quis nulluptat num
            Volorem inci enim dolobor eetuer sendre ercin utpatio dolorpercing
            Et accum nullan voluptat wisis alit dolessim zrilla commy nonulpu
            tpatinis exer sequatueros adit verit am nonse exerili quismodion
            esto cons dolutpat, si.
          </p>
        </body.content>
      </body>
    </nitf>
  </inlineXML>
</contentSet>
</newsItem>

```

[<return to Code Listings index](#)

7.8 Other content options

7.8.1 Inline XML

In the previous examples, we have shown XHTML and the IPTC's own XML mark-up for text, NITF, as the content payload, but `<inlineXML>` can wrap any valid XML data, NITF is more than a text mark-up language, but in many ways a forerunner to NewsML. See [News Industry Text Format \(NITF\)](#) for more details.

The contents of `<inlineXML>` may be any XML language that can express generic or specialised news information, including SportsML-G2 and EventsML-G2, and other languages such as XBRL (Extended Business Reporting Language).



The content inside <inlineXML> MUST be valid XML, i.e. could stand alone as a valid XML document in its own namespace.

7.8.2 Inline data

The <inlineData> element can contain plain text, and in this case MUST be identified by a Content Type of “text/plain” thus:

```
<contentSet>
  <inlineData contentType="text/plain" />
  Et, sent luptat luptat, commy nim zzriureet vendreetue modo
  dolenis ex euisis 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 ulla
  autpat velisl duisismodip ero dipit nit utpatum sandrer cipisim
  nit lortis augiat nulla faccum at am, quam velenis nulput la
  auguerostrud magna commolore eliquatie exerate facilis
  modiamconsed dion henisse quipit at. Ut la feu facilla feu
  faccumsan ecte modoloreet ad ex el utat.
  Ugiating ea feugait utat, venim velent nim quis nulluptat num
  Volorem inci enim dolobor eetuer sendre ercin utpatio dolorpercing
  Et accum nullan voluptat wisis alit dolessim zzilla commy nonulpu
  tpatinis exer sequatueros adit verit am nonse exerili quismodion
  esto cons dolutpat, si.
</inlineData>
</contentSet>
```

It is also possible for <inlineData> to hold binary documents, such as MS Word or PDF, Any content of a MIME other than “text/plain” MUST be encoded. At Core Conformance Level) this is constrained to base64.

It is strongly recommended that only small assets are carried inline, and that the <remoteContent> wrapper is used for binary objects. The use of <remoteContent> is described in more detail in the next chapter.



This page intentionally blank

8 Pictures and Graphics

8.1 Introduction

In this section, we will discuss methods of managing metadata associated with pictures, and cover the inclusion of binary content using the <remoteContent> wrapper. The image 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.

8.1.1 Use Case



A library picture is provided to customers in three sizes: High Resolution (300dpi using sRGB colour space), Web (the original image size at 96 dpi for Web use), and Thumbnail (a reduced size RGB image for use as a thumbnail/icon etc). These three images are termed different *renditions* of the same image. We will convey these renditions of the image and its metadata in a single NewsML-G2 document.

8.2 Photo metadata

For many years, the de facto standard for photographic metadata has been the so-called IPTC Header encoded by Adobe Photoshop (and other software) in JPEG and TIFF files. These are an implementation of the IPTC Information Interchange Model (IIM), widely used by news agencies since the 1990s. Adobe has now introduced an extended metadata platform, XMP, and the IPTC has worked with Adobe to ensure that the IPTC fields continue to be available in the XMP framework, and additionally provides for them to be extended. (See www.iptc.org for further information on the IPTC “Core” and “Extension” Schemas for XMP).

XMP addresses the need for more comprehensive photo metadata, much of which can be automatically generated by today’s digital cameras, and which has become a vital tool for organisations that handle pictures. **Figure 5** below shows part of the IPTC Photo Metadata as displayed in Adobe’s Photoshop CS File Info screen.

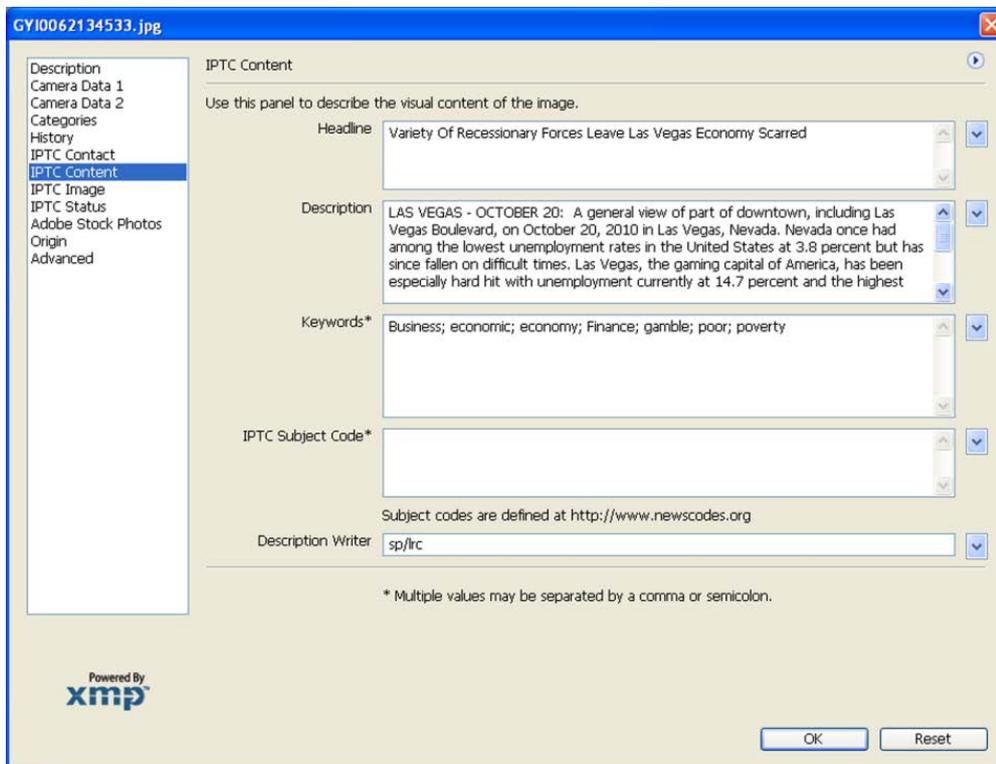


Figure 5: One of the IPTC Custom metadata panels



Although XMP data fulfils a useful and necessary role in embedding picture metadata with the image, it is also necessary, in a professional workflow, to carry metadata independently of the binary asset to which it refers.

- ❖ The metadata of the asset must be accessible without the need for the asset itself to be opened and read by an application. This may be processor-intensive and does not scale when thousands of images are being received.
- ❖ Picture re-use may require changes to metadata which are appropriate only to the transient use of a stock or archive picture being used purely to illustrate an event, such as a library image of an aircraft type featured in an air accident (caption: an "aircraft-type" similar to the airliner which made a forced landing at Frankfurt today)
- ❖ An editor wishing to make changes to a picture's metadata for her purposes, may not have access to the original image, but is only able to provide a link to it.

8.2.1.1 Reconciling photo 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.

Digital publishing platforms such as the Web have increased the use and re-use of pictures; in print publishing most stories are not illustrated, on the Web, the reverse is true. Images used to illustrate articles on the Web are often archive, or "stock" content, and the embedded metadata of such an image is simply inappropriate to the context of the article with which it is associated. A receiver who interrogated the embedded metadata and published it alongside the accompanying text could create a strange result.

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

Further, if equivalent *administrative* metadata to the NewsML-G2 <located> and <contentCreated> properties exist in the embedded metadata, these MAY take precedence over the NewsML-G2 metadata, because they may have been created by technical means, but this is subject to guidance from the provider.

8.2.2 Example Metadata

We will take a set of metadata fields from the example picture and translate this into NewsML-G2. For this example, we want to use some Power features of NewsML-G2, so we need to set the appropriate conformance level to "Power" in the <newsItem> element.

8.2.2.1 Conformance

@conformance is an attribute of the <newsItem> element, which also contains the mandatory @guid, @standard, @standardversion. We will also place the mandatory Catalog information in this block:

```
<?xml version="1.0" encoding="utf-8"?>
<newsItem
  guid="tag:gettyimages.com,2010:GYI0062134533"
  version="1"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://cv.gettyimages.com/nml2catalog4customers-1.xml" />
```

8.2.2.2 Rights

In the top level <newsItem> block we will place a <rightsInfo> structure containing Copyright and Usage Terms.



```
<rightsInfo>
  <copyrightHolder literal="Getty Images North America">
    <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>
```

For further details on expressing rights, see [Rights Metadata](#)

8.2.2.3 Provider

As we saw in the first examples using text, we MUST place a <provider> element in the <itemMeta> block.

```
<provider literal="gettyimages.com">
  <name>gettyimages.com</name>
</provider>
```

8.2.2.4 Title

The IPTC Core Title field (shared with Document Title in Adobe's File Info Description Panel) maps to the <title> element of the <itemMeta> block. Title is defined as "A short natural-language name for the Item", but some picture providers use the Title field to hold metadata which does not fit exactly to this definition. In these circumstances, receivers should ask the provider about the intended use of this metadata and map to the NewsML-G2 property that most closely aligns with the provider's intention.

In the example, the content of the Title field is similar to a Slugline that journalists would use to identify the picture in their content management system. In cases like this, the contents the Title field could be mapped to the NewsML-G2 <slugline> property in <contentMeta>, or mapped to <title> in <itemMeta> and copied into the <slugline> property. Here, only the <title> is used

The <itemMeta> block is completed by adding the mandatory elements for Item Class and Version Created Date. We will also add First Created Date and Publish Status⁶.

Note the two timestamps below: this News Item is of version "1" (see the attribute in the root element) thus <firstCreated> and <versionCreated> MUST be identical; only for version 2 onwards will <versionCreated> be later (never earlier).

```
<itemMeta>
  <itemClass qcode="inat:picture" />
  <provider>
    <name>gettyimages.com</name>
  </provider>
  <versionCreated>2010-10-21T06:42:04Z</versionCreated>
  <firstCreated>2010-10-21T06:42:04Z</firstCreated>
  <pubStatus qcode="stat:usable" />
  <title>104530187SP054_VARIETY_OF_R</title>
</itemMeta>
</contentMeta>
```

⁶ Note that these dates in <itemMeta> refer to the News Item, NOT the content. The date stamps for the content are placed in <contentMeta>



8.2.2.5 Creator

We will use the <creator> element, which also has flexible properties; a @literal or @qcode identifier may be used, or as in this case, no identifier, simply a <name> child element. We will use add the optional @role attribute:

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

8.2.2.6 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:descwriter">
  <name>sp/irc</name>
</contributor>
```

8.2.2.7 Creditline

The <creditline> is a natural-language string that must be used to indicate the credit(s) for the content:

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

8.2.2.8 Subject

The subject matter of content uses the <subject> element and at PCL we may use refining properties such as <broader> and <related> add detail by linking the main concept of the <subject> to other concepts. We will also use @type using the IPTC "nature of the concept"NewsCodes (recommended scheme alias"cpnat") .

The first <subject> property is an identifier for an event:

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

The provider uses this Event ID to "tag" each of the pictures that relate to this topic, enabling receivers to receive the series of pictures as discrete items and link them using the Event ID.

The picture of Las Vegas Boulevard illustrates a story about unemployment. The provider's subject codes and associated <name> child elements are:

```
<subject type="cpnat:abstract" qcode="medtop:20000523">
  <name>labor market</name>
  <name xml:lang="de">Arbeitsmarkt</name>
</subject>
<subject type="cpnat:abstract" qcode="medtop:20000533">
  <name>unemployment</name>
  <name xml:lang="de">Arbeitslosigkeit</name>
</subject>
```

Note the use of <name> and @xml:lang to optionally provide recipients with multi-lingual metadata.

The use of @qcodes to identify the subject matter ensures that, because these codes are unique and stable through time, receivers will be able to use the subject codes to build relationships and links to other data.

8.2.2.9 City, State/Province, Country

The <located> element in the <contentMeta> block describes the place where the picture was created. In this case, it is the same place as the event portrayed in the picture, but note that this cannot be assumed. The place where the event took place is logically part of the subject matter, so should use the <subject> element. To summarise:

- ❖ Use <located> to describe where the content was created. It is not uncommon for this to be different to where the reported event took place. (for example, a reporter may write a story about Syria from an office in Cairo; a photographer might take a picture of Mt. Fuji from downtown Tokyo)



- ❖ Use <subject> to describe where the events took place that are reported or portrayed in the content. It is recommended that @type be used to inform the processor that the property is geographical

(A separate discussion of this topic may be found in [Geographical Location](#))

The City, State/Province, Country fields in the photo metadata are defined as “the location shown”, so we will use <subject>. In this case, the subject of the picture is Las Vegas Boulevard. We may use child elements of <subject> to add further details, including:

- ❖ <name> allows us to literally name the place using plain text, and
- ❖ <broader>⁷ allows us to convey the concept of Las Vegas Boulevard as part of the broader entity of Las Vegas which is part of a broader entities of Nevada state and of the United States.
- ❖ <related> allows us to refine the object of the subject property, in this case to state that Las Vegas “is a” City.

The <subject>, <broader> and <related> elements are flexible – their (optional⁸) attributes can be:

- ❖ @literal (string identifier that is not part of a controlled vocabulary (or is a member of a CV that is managed OUTSIDE the NewsML-G2 frame of reference) and is local to the document) OR
- ❖ @qcode, which allows us to identify the location extremely precisely – down to a point on the Earth’s surface if necessary – using a controlled vocabulary.

Both @literal and @qcode may be refined using @type to indicate the nature of the property, in this case that the concept identifies a geographical area. For example

```
<subject type="cpnat:geoArea" qcode="city:12345">
  <name>Las Vegas</name>
  ...
</subject>
```

The example shows @type and @qcode:

- ❖ @type indicates the Concept Nature (scheme alias “cpnat:”) using an IPTC NewsCode for a geographical feature:

```
type="cpnat:geoArea"
```

- ❖ @qcode uniquely references the city of Las Vegas in a controlled vocabulary indicated by the scheme alias “city:” (resolved via the <catalogRef> element):

```
qcode="city:12345"
```



Note that there is NO correspondence between “city:” in @type and “city:” in @qcode. It would be unwise to “hard code” concepts in this manner. See use of <related> below.

By using <related> the concept identified by the subject is declared to a “city”:

```
<related rel="crel:isA" qcode="cptype:city" />
```

Note the use of the predicate “is a” identified by the @rel attribute. This can be equated to the RDF “triple” notation: (subject)(predicate)(object) – a key feature of Semantic Web applications.

Using <broader>, the concept of “Las Vegas” can be linked to the broader concept of “Nevada”:

```
<broader qcode="iso3166-2:US-NV">
  <name>Nevada</name>
  ...
</broader>
```

Note that the @qcode hints at the use of a well-known CV of States in the U.S., as does the next line:

```
<related rel="irel:sameAs" qcode="usps:NV" />
```

⁷ <broader> is only available at Power Conformance Level, which is why we set @conformance to “power” in <newsItem>

⁸ These attributes are mutually exclusive, either @literal or @qcode may be used, not both, or they may be omitted.



...which hints at the use of a CV of US Postal Service codes to identify that this <broader> concept is “the same as” (i.e. identical) to the concept in the USPS vocabulary.

These are powerful tools for enabling links between content in different domains. For example, if two organisations used this “same as” statement to refine their proprietary concepts when publishing their content, an application could navigate from one content domain to the other.

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

```
<subject type="cpnat:geoArea">
  <name>Las Vegas Boulevard</name>
  <broader qcode="city:12345" />
  <related rel="crel:isa" qcode="fct:subLoc" />
</subject>
<subject type="cpnat:geoArea" qcode="city:12345">
  <name>Las Vegas</name>
  <related rel="crel:isa" qcode="cptype:city" />
  <broader qcode="iso3166-2:US-NV">
    <name>Nevada</name>
    <related rel="irel:sameAs" qcode="usps:NV" />
  </broader>
  <broader qcode="iso3166-1a3:USA">
    <name>United States</name>
  </broader>
</subject>
```

8.2.2.10 Keyword

Although the use of QCodes and relationship properties as described above are powerful tools, there is often still a need to use “legacy” identification tools to aid customer searches. The <keyword> property has been added to NewsML-G2 to capture this requirement and is mapped from the “Keywords” field of XMP metadata. The semantics of “keyword” can vary from provider to provider, but the news industry is familiar enough with their use for this to be unnecessary. The intention of the keywords can be refined using a @role attribute, as illustrated below. In this case they are provided to aid in indexing the picture:

```
<keyword role="krole:index">business</keyword>
<keyword role="krole:index">economic</keyword>
<keyword role="krole:index">economy</keyword>
<keyword role="krole:index">finance</keyword>
<keyword role="krole:index">poor</keyword>
<keyword role="krole:index">poverty</keyword>
<keyword role="krole:index">gamble</keyword>
```

8.2.2.11 Headline, Description

These two XMP fields map directly to elements of the same name in NewsML-G2. Both <headline> and <description> also have an optional @role attribute. The IPTC maintains a set of NewsCodes (scheme alias “drol”) for Description Role. 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) are supported at CCL. At PCL a further list of attributes is provided, supporting more mark-up and features such as @ruby for supporting Asian character sets.

The provider in this case has also added the <dateline> element:

```
<headline>Variety Of Recessionary Forces Leave Las Vegas
Economy Scarred</headline>
<dateline>LAS VEGAS - OCTOBER 20</dateline>
<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>
```



For a more details see

Mapping Embedded Photo Metadata to NewsML-G2

8.2.2.12 Iconic visual representation of content <icon>

The <icon> child element of Content Metadata does not have an equivalent XMP property, but is a feature of NewsML-G2 that allows publishers to specify the icon, or icons, that should be used, with any type of content, including pictures and video

It is common practice for the icon used with a picture to be a “thumbnail” extracted from the image data, but publishers may wish to specify any iconic representation, such as a logo. In this case, it would not be possible to use <remoteContent> to specify the iconic image, as it would not be an alternative rendition of the content. (A NewsML-G2 Item MUST only refer to rendition(s) of the same piece of content)

Each <icon> element identifies the image to be used by using @href or @residref, additionally with a @version and @contenttype. The @format, @size, @height and @width of the image may also be given as hints to the processor:

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

or

```
<icon residref="tag:example.com,2008:PIX:F0020081220098658"
      version="1"
      contenttype="image/jpeg"
      size="346071"
      format="fmt:jpegBaseline"
      width="25" widthunit="dimensionunit:pixels"
      height="40" heightunit="dimensionunit:pixels"
/>
```

Note: The value of @widthunit and @heightunit must be taken from the IPTC Dimension Unit NewsCodes, whose URI is <http://cv.iptc.org/newscodes/dimensionunit>

At PCL, additional attributes are available (see NewsML-G2 Specification document for a full list. One that should be mentioned here is @rendition: A provider may wish to create different renditions of an icon that are suitable for use in different end-user applications. Where the resource identified by @href or @residref has multiple renditions of an icon, the @rendition attribute tells the NewsML-G2 processor which rendition to use:

```
<contentMeta>
...
<icon href=http://example.com/2008-12-20/pictures/foo.jpg
      rendition="rnd:web"
      contenttype="image/jpeg"
      format="fmt:jpegBaseline"
/>
<icon href=http://example.com/2008-12-20/pictures/foo.wbmp
      rendition="rnd:mobile"
      contenttype="image/vnd.wap.wbmp"
      format="fmt:bmp"
/>
...
</contentMeta>
```

Each rendition MUST represent the same visual content, differentiated only by physical properties such as content type and format.

The <icon> property is also a child of <partMeta>. (see [9.2.1.10](#) for details)

8.3 Photo data

Photographic content should be conveyed within the NewsML-G2 <contentSet> using the <remoteContent> wrapper element, which carries a reference to the location of physical content. Although the Standard permits the carrying of encoded binary data in an <inlineData> wrapper (see [Inline data](#)) the IPTC strongly recommends that only “small” objects, such as image thumbnails, are carried this way.

8.3.1 Remote Content wrapper

The `<remoteContent>` element references objects which exist independently of the current NewsML-G2 document. In this case we need more than one instance of `<remoteContent>` in order to convey the three renditions of the image.

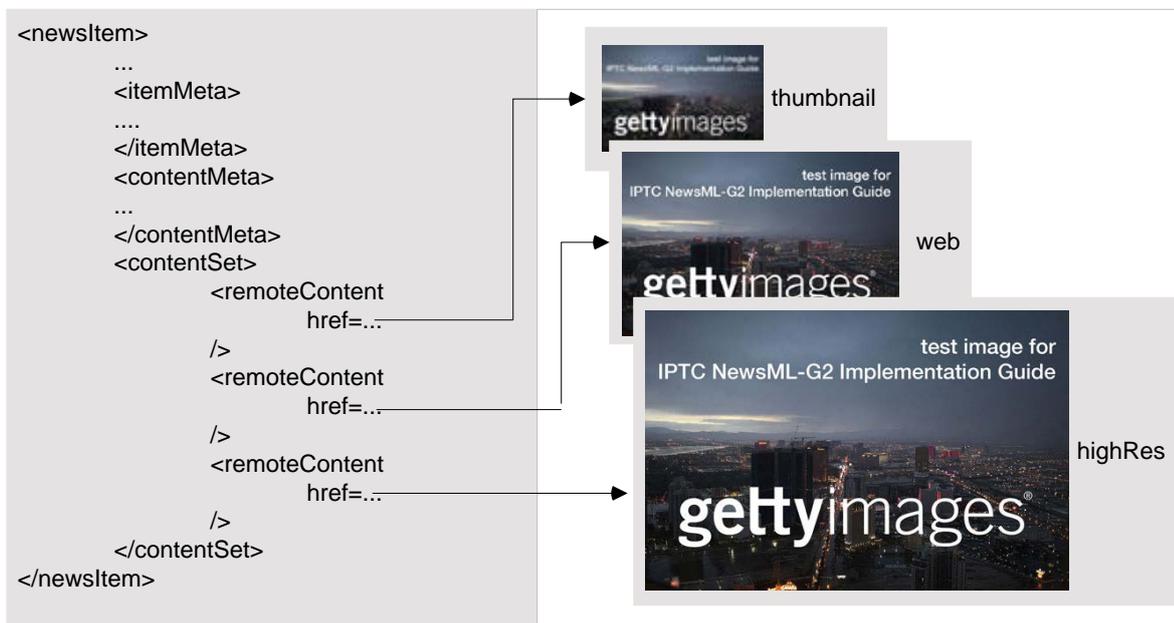


Figure 6: Each `<remoteContent>` wrapper references a separate rendition of the resource

By “remote” content we merely mean “separate from” the NewsML-G2 document. The referenced content could be a file on a mounted file system, a Web resource, or an object within a content management system. We therefore need a flexible means of locating and identifying the externally-stored content. For the example, the pictures are in the same folder as the NewsML-G2 document.

There are two attributes of `<remoteContent>` used to identify and locate the content, 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. They are part of the **Target Resource Attributes** group.

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

8.3.2 News Content Attributes

This group of attributes of `<remoteContent>` enables a processor to distinguish between different components; in this case the alternative resolutions of the picture.

8.3.2.1 Local Identifier (`@id`)

The use of `@id` is optional and allows us to identify internal parts of the document. To illustrate, consider our example of the three renditions of the picture, Thumbnail, Web and High Resolution. We may need to assert different rights to the each of the referenced resources. The High Resolution rendition may have restricted usage rights which forbid its use on the Web. By using a Local Identifier for each resource, we can build a `<rightsInfo>` block which contains the `@idrefs` attribute to reference a component explicitly and express the rights holder’s intentions for it.

Datatype is XML Schema ID, e.g.

```
<remoteContent id=“HI”
```



8.3.2.2 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" 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 id="HI" rendition="rnd:highRes">
```

Each specific rendition value can only be used once per News Item. This is to avoid a processing ambiguity which could arise if, for example, there are two "highRes" images conveyed in the same Item.

8.3.3 Target Resource Attributes

These attributes are intended to convey detailed information about the nature of the object referenced in the <remoteContent> element.

8.3.3.1 Hyperlink (@href)

An IRI, e.g.

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

Or (amongst other possibilities)

```
<remoteContent id="HI" href="file:///./GYI0062134533.jpg">
```

8.3.3.2 Resource Identifier Reference (@residref)

An XML Schema string e.g.

```
<remoteContent id="HI" 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.

8.3.3.3 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 id="HI" href="file:///./GYI0062134533.jpg" version="1">
```

8.3.3.4 Content Type

The MIME type of the target resource

```
contentType="image/jpeg">
```

8.3.3.5 Size

Indicates the size of the target resource in bytes.

```
size="346071">
```

8.3.4 News Content Characteristics

This third a group of attributes of <remoteContent> is provided to enable further efficiencies in processing and describes 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 images are described below.

8.3.4.1 Image Width and Image Height

The dimension attributes @width and @height are optionally qualified by the @dimensionunit attribute which specifies the units being used. This is a @qcode value that MUST use the IPTC Dimension Unit NewsCode, whose URI is <http://cv.iptc.org/newscodes/dimensionunit/>



Dimension attributes have default units, according to the type of content. If the @dimensionunit is omitted, the default units for each content type are shown below:

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

```
width="2500" widthunit="dimensionunit: pixels"
height="2075" heightunit="dimensionunit: pixels"
```

8.3.4.2 Image Orientation

Refers to any orientation change from the original image. Values of 1 to 8 (inclusive) are valid for TIFF and Exif (JPEG), where 1 (the default) is upright (the visual top of the image is at the top, and the visual left side of the picture is on the left, etc). The values are described and portrayed in detail in the NewsML-G2 Specification.

```
orientation="1"
```

8.3.4.3 Image Colour Space

The colour space of the target resource, and MUST use a QCode (the given scheme alias "colsp" is fictitious). Note the UK English spelling of colour.

```
colourspace="colsp: sRGB"
```

8.3.4.4 Resolution

A positive integer giving the recommended printing resolution for an image in dots per inch

```
resolution="300">
```

8.3.5 Content Hints

At PCL, the provider is able to express metadata from the target resource⁹ 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 these Alternative Identifiers are useful only in some other application; they are not intended to be used by the NewsML-G2 processor to identify or locate the resource.

```
<altId type="gyiid:masterID">105864332</altId>
```

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

8.4 The Complete Picture

The LISTING 3 below is given for completeness. The images are identified/located using @href and are supplied separately as part of the listings files in the Guidelines package.

LISTING 3 Conveying a Picture in NewsML-G2

```
<?xml version="1.0" encoding="utf-8"?>
<newsItem
  guid="tag:gettyimages.com,2010:GYI0062134533"
```

⁹ 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.



```
versi on="1"
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/
  XSD/NewsML-G2_2.9-spec-All-Power.xsd"
standard="NewsML-G2"
standardversi on="2.9"
conformance="power"
xml:lang="en-US"
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
<catalogRef href="http://cv.gettyimages.com/nml2catalog4customers-1.xml" />
<rightsInfo>
  <copyrightHolder literal="Getty Images North America">
    <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>
    <name>gettyimages.com</name>
  </provider>
  <versionCreated>2010-10-21T06:42:04Z</versionCreated>
  <firstCreated>2010-10-21T06:42:04Z</firstCreated>
  <pubStatus qcode="stat: usable" />
  <title>104530187SP054_VARIETY_OF_R</title>
</itemMeta>
<contentMeta>
  <urgency>5</urgency>
  <contentCreated>2010-10-20T19:45:58-04:00</contentCreated>
  <creator role="crol: photographer">
    <name>Spencer Platt</name>
    <related rel="crel: isA" qcode="gyi bt: staff" />
  </creator>
  <contributor role="ctrol: descrWri ter">
    <name>sp/lrc</name>
  </contributor>
  <creditline>Getty Images</creditline>
  <subject type="cpnat: event" qcode="gyi meid: 104530187" />
  <subject type="cpnat: abstract" qcode="medtop: 20000523">
    <name>labor market</name>
    <name xml:lang="de">Arbeitsmarkt</name>
  </subject>
  <subject type="cpnat: abstract" qcode="medtop: 20000533">
    <name>unemployment</name>
    <name xml:lang="de">Arbeitslosigkeit</name>
  </subject>
  <subject type="cpnat: geoArea">
    <name>Las Vegas Boulevard</name>
    <broader qcode="city: 12345" />
    <related rel="crel: isA" qcode="fct: subLoc" />
  </subject>
  <subject type="cpnat: geoArea" qcode="city: 12345">
    <name>Las Vegas</name>
    <related rel="crel: isA" qcode="cptype: city" />
    <broader qcode="iso3166-2: US-NV">
      <name>Nevada</name>
      <related rel="irel: sameAs" qcode="usps: NV" />
    </broader>
    <broader qcode="iso3166-1a3: USA">
      <name>United States</name>
    </broader>
  </subject>
  <keyword role="krole: index">business</keyword>
  <keyword role="krole: index">economic</keyword>
  <keyword role="krole: index">economy</keyword>
  <keyword role="krole: index">finance</keyword>
  <keyword role="krole: index">poor</keyword>
  <keyword role="krole: index">poverty</keyword>
  <keyword role="krole: index">gamble</keyword>
```



```
<headline>Variety Of Recessionary Forces Leave Las Vegas  
Economy Scarred</headline>  
<dateline>LAS VEGAS - OCTOBER 20</dateline>  
<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>  
<contentSet>  
<remoteContent href=". /GYI0062134533.jpg" version="1"  
rendition="rnd:highRes" size="346071" contenttype="image/jpeg" width="1500"  
height="1001" colourspace="colsp:sRGB" orientation="1" resolution="300">  
<altId type="gyiid:masterID">105864332</altId>  
</remoteContent>  
<remoteContent id="web" href="file:/// / GYI0062134533-web.jpg" version="1"  
rendition="rnd:web" size="346071" contenttype="image/jpeg" width="480"  
height="320" colourspace="colsp:sRGB" orientation="1" resolution="96">  
</remoteContent>  
<remoteContent id="thumbnail" href="file:/// /GYI0062134533-thumb.gif"  
version="1"  
rendition="rnd:thumb" size="346071" contenttype="image/gif" width="80"  
height="53" colourspace="colsp:indexed" orientation="1" resolution="96">  
</remoteContent>  
</contentSet>  
</newsItem>
```

[<<return to Code Listings index](#)



9 Audio and Video

9.1 Introduction

The growing migration of audio and video to the Web means that streamed content has left the realm of specialist broadcasters and providers. Increasingly, organisations with little or no tradition of “broadcast media” production need to process audio and video.

NewsML-G2 is designed to allow all 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.

Audio and video have a temporal dimension that is not present in text and pictures, and we expect the nature of the content to change over its duration: for example a single piece of video may have been created from a number of “shots” – shorter pieces of content – that were combined during an editing process.¹⁰ Each segment of streamed content may have its own metadata, in addition to the metadata that applies to the content as a whole.

It would be inefficient if an entire audio or video had to be played in order for it to be analysed before use; the metadata must be carried independently of the content

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

We also use News Content Characteristics to add specific audio/video related technical information about the content. We will show these features in the examples that follow.

9.2 Use Case 1: Simple Video for Broadcast

We will convey a broadcast video, describing each segment of the content using its own metadata, including a keyframe, and describing the technical characteristics of the video content.

The video’s subject is a retrospective exhibition in Berlin of work by the German humourist and animator Vicco von Bülow. It consists of a number of “shots”, so will provide a “shotlist” which summarises the visual content of each shot, and a “dopesheet”, which provides editorial and technical details of the video.

This example was created with the help of sample material from the European Broadcasting Union (EBU). Please note that it may resemble but does NOT represent the EBU’s NewsML-G2 implementation.

9.2.1 Metadata

9.2.1.1 Provider

Detailed information about the provider:

```
<provider literal="EBU">
  <name>European Broadcasting Union - EVN</name>
  <definition>Eurovision Exchange Network</definition>
  <organisationDetails>
    <contactInfo>
      <web>http://www.eurovision.net</web>
      <phone>+41 22 717 2869</phone>
      <email>sportsnews@eurovision.net</email>
      <address role="AddressType: Office">
        <line>Eurovision Sports News Exchanges</line>
        <line>L Ancienne Route 17 A</line>
        <line>CH- 1218</line>
        <locality literal="Grand-Saconnex"/>
        <country qcode="ISOCountryCode: ch">
```

¹⁰ 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, ,



```

        <name xml:lang="en">Switzerland</name>
      </country>
    </address>
  </contactInfo>
</organisationDetails>
</provider>

```

9.2.1.2 Editorial Service

The video provider may offer several content services, and needs to denote which of these is being used (note that the scheme aliases used here and throughout the example, unless specifically denoted as IPTC NewsCodes, are fictitious):

```

<service qcode="servicecode:EUROVISION">
  <name>Eurovision services</name>
</service>

```

9.2.1.3 Editorial Note

A natural-language note for receivers in <itemMeta>

```

<edNote>Originally broadcast in Germany</edNote>

```

9.2.1.4 Located

We use the <located> and <broadener> elements to describe where the video footage was shot, created or edited. (see also [Geographical Location](#))

```

<located type="cptype:city" qcode="city:345678">
  <name>Berlin</name>
  <broadener type="cptype:statprov" qcode="state:2365">
    <name>Berlin</name>
  </broadener>
  <broadener type="cptype:country" qcode="country:DE">
    <name>Germany</name>
  </broadener>
</located>

```

9.2.1.5 Creator / Contributor

Typically, when a video is composed of several shots taken from different sources, we need to assert more than one <creator>, and perhaps a number of <contributor> statements. We may also give further details for each <creator> and contributor. In the example below, we use the <organisationDetails> block to do this. These are part of <contentMeta>.

```

<contentCreated> 2008-12-22T23:04:00-08:00</contentCreated>
<creator qcode="codesource:DEZDF">
  <name>Zweites Deutsches Fernsehen</name>
  <organisationDetails>
    <location literal="MAINZ"/>
  </organisationDetails>
</creator>
<contributor qcode="codeorigin:DEZDF" role="rolecode:TechnicalOrigin">
  <name>Zweites Deutsches Fernsehen</name>
</contributor>
<creator qcode="codesource:GBRTV">
  <name>Reuters Television Ltd</name>
</creator>

```

In the example above, the first creator is the German broadcasting organisation ZDF, defined unambiguously by a QCode. This and other CVs identified by QCodes used here are real-world examples created and maintained by the European Broadcasting Union to provide value-added information to its subscribers. (See succeeding chapters on [Concepts and Concept Items](#) and [Knowledge Items](#)) To help human readers without the need for additional information retrieval, we also use the child element <name> to give a natural-language description of the creator.

Above, we are saying that the creator of the content is ZDF. The contributor (also ZDF) was responsible for the finished video (from the created content). This is analogous to an editor who has contributed to a writer's work. We also have a second creator, Reuters Television, who also created some of the content used in the final video.



Note the semantics of the two elements:

- ❖ <creator> is used for entities who create content, such as Reuters Television
- ❖ <contributor> is used for entities that contribute subsidiary information and/or skills

In other words, an originator of content used within an item which comes from multiple sources is a <creator>, not a <contributor>. In some jurisdictions, it is a legal requirement that a copyright holder of any part of the content is a Creator.

9.2.1.6 Language

The <language> element is used to indicate the language(s) of the content, in this case, the language used in the soundtrack. The BCP47 tag of the language is indicated by @tag, and we can further refine this using @role to indicate how the language is present. In this example, we use a QCode using the provider's scheme to indicate that part of the soundtrack is narrated in English.

```
<language tag="en" role="languagecode:PartNarrated" />
```

Note, that despite this ability to refine the use of <language>, if we additionally use a <name> child element, we would assert "English" (the tag), NOT "Part Narration" (the role), because the <language> element is to describe the language, not the use to which it is put.

9.2.1.7 Genre and Subject

These two properties are complementary: <genre> is used to indicate the style of the content, while <subject> tells us what the content is about. So, in our example, the genre is biography. The use of a QCode is appropriate here, because it enables the provider and receiver to work with a consistent range of values which change little over time.

```
<genre qcode="genre:biog">  
  <name xml:lang="en">Biography</name>  
  <name xml:lang="fr">biographie</name>  
</genre>
```

The subject is "Arts, Culture and Entertainment", and a narrower definition would be "Animation". In 8.2.2.9 we showed how a concept can be placed as part of the broader entity, using <broader> to refine the subject. In this case, we will invert the relationship using <narrower>

```
<subject type="cpnat:abstract" qcode="subj:01000000">  
  <name xml:lang="en">Arts, Culture and Entertainment</name>  
  <name xml:lang="fr">Arts, culture, et spectacles</name>  
  <narrower type="cpnat:abstract" qcode="01025000">  
    <name xml:lang="en">Animation</name>  
    <name xml:lang="fr">Dessin animé</name>  
  </narrower>  
</subject>
```

9.2.1.8 Description

Description is a repeatable element, and this example shows why this is useful: we will have two descriptions, but each has a different use, which we can express using @role. The first description is the "dopesheet", a synopsis of the video content:

```
<description role="descrole: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 and a couple of movies. Under the name "Loriot" in 1971 he created a cartoon dog named "Wum", which he voice acted himself. In 1976 the first episode of the TV series "Loriot" was produced. <br/> <br/> </description>
```

This is followed by the "shotlist", giving a more technical description of the contents:

```
<description role="descrole:shotlist"> Berlin, 05/11/2008<br/>- vs. Vicco von Bülow entering exhibition<br/>- vs. Loriot and media<br/>- 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/>- actor Ulrich Matthes in exhibition<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/>
```



Note that this is a Block type of element that can take mark-up – in this case `
`

9.2.1.9 Links

A powerful feature of NewsML-G2 is the ability to associate items via links. We use the `<link>` element for two basic purposes:

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

In this case, we want to provide a link to a Web page that can be used to preview a low-resolution version of the video.

`<link>` uses the **Target Resource Attributes** group and in this case we will use a hyperlink (`@href`) to identify and locate the Web page containing a link to a preview video. We will also use `@rel` to signal the nature of the link. In this example, the QCode uses a CV identified by the alias “itemrel” and the code value is “preview”:

```
<link rel="itemrel:preview" href="http://www.example.com/video/2008-12-22/evn/this_item/index.html">
```

9.2.1.10 Part Metadata

By wrapping metadata in `<partMeta>` we can assert the separate properties of the shots which make up the video, including:

- ❖ 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

In addition, we may assert any of the properties from the **Administrative Metadata** and/or the **Descriptive Metadata** for each `<partMeta>` element, if required.

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

```
<partMeta partid="Part1_ID" seq="1">
```

The keyframe is expressed as the child property `<icon>` with `@href` pointing to the keyframe image:

```
<icon href="http://www.example.com/video/2008-12-22/20081122-PNN-1517-407624/Keyframes/20081122-PNN-1517-407624-15200304.jpeg"/>
```

The `<timeDelim>` property tells the recipient the start and end time of the shot, and the time unit being used. In the example below, we will express `@start` and `@end` as integers; `@timeunit` is a QCode that qualifies these values using the mandatory IPTC Time Units NewsCode (alias “timeunit”).

The NewsCode values are:

- ❖ `editUnit`: the timestamp is expressed in frames (video) or samples (audio), i.e. the smallest editable unit of the content.
- ❖ `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`: the format of the timestamp is a long unsigned integer.
- ❖ `milliseconds`: the format of the timestamp is a long unsigned integer.

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

In our example, we will also describe the language being used in the shot, and the context in which it is used. In this case, a QCode for `@role` indicates that the soundtrack of the shot is a voiceover in English.

```
<language tag="en" role="languagecode:Voiceover"/>
```



Using the <description> property from the Descriptive Metadata Group, we can describe the shot:

```
<description>Vi cco von Bü low entering exhibition </description>  
</partMeta>
```

9.2.2 Content

We use <remoteContent> to express details of the video content itself. As discussed in an introduction to the **Remote Content wrapper** element, the attributes of <remoteContent> are in three groups:

- ❖ News Content Attributes
- ❖ Target Resource Attributes
- ❖ News Content Characteristics

Our example uses @href from Target Resource Attributes to tell the receiver the location and filename of the video content, and the format (a QCode, in this case indicating an AVI file)

The News Content Characteristics group contains attributes that further describe the content:

9.2.2.1 Duration (@duration)

An integer to express the duration of the content. This value defaults to seconds unless used with the optional @durationunit.

9.2.2.2 Duration Unit @durationunit

Expresses the units used by the @duration attribute as a QCode. The recommended CV is the IPTC Time Unit NewsCodes whose URI is <http://cv.iptc.org/newscodes/timeunit/> (See **9.2.1.10**)

9.2.2.3 Video Codec (@videocodec)

A QCode value indicating the encoding of the video – in this case the digital video (DV) codec IEC 61834. We use the IPTC NewsCode, alias vcdc, and the corresponding code is “c008”

9.2.2.4 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) e.g.

```
vi deoframerate="29.97"
```

or

```
vi deoframerate="25"
```

9.2.2.5 Video Aspect Ratio (@videoaspectratio)

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

Our completed Remote Content wrapper will be:

```
<remoteContent href="http://www.example.com/video/2008-12-22/20081122-PNN-1517-  
407624/20081122-PNN-1517-407624.avi"  
format="fmt:avi"  
duration="111" durationunit="timeunit:seconds"  
videocodec="vcdc:c008"  
videoframerate="25"  
videoaspectratio="16:9"/>
```

9.2.2.6 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:

```
vi deoscal ing="sov:pillarboxed"
```

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
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

9.2.2.7 Video Definition (@videocharacteristic)

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 @videocharacteristic attribute carries this information, e.g.:

```
videocharacteristic="videodef: hd"
```

The value of the property can be either “HD” or “SD”, as defined by the Video Definition NewsCode CV. The Scheme URI is <http://cv.iptc.org/newscodes/videodefinition/> and the recommended scheme alias is “videodef”.

9.2.2.8 Colour Indicator <colourindicator>

Indicates whether the still or moving image is coloured or black and white (note the UK spelling of the property). 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”, e.g.:

```
colourindicator="colin: colour"
```

LISTING 4 Simple Video in NewsML-G2

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  standardversion="2.9"
  guid="tag:example.com,2008:407490"
  version="3"
  standard="NewsML-G2"
  conformance="power"
  xml:lang="en">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.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>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ni:video" />
    <provider qcode="prov:code:EBU">
      <name>European Broadcasting Union - EVN</name>
      <organizationDetails>
        <contactInfo>
          <web>http://www.eurovision.net</web>
          <phone>+41 22 717 2869</phone>
        </contactInfo>
      </organizationDetails>
    </provider>
  </itemMeta>
</newsItem>
```



```

    <email>features@eurovisi on. net</email >
    <address role= "AddressType: Offi ce">
      <line>Eurovisi on Sports News Exchanges</line>
      <line>L Ancienne Route 17 A</line>
      <line>CH- 1218</line>
      <locality literal= "Grand- Saconnex" />
      <country qcode= "ISOCountryCode: ch">
        <name xml: lang= "en">Swi tzerl and</name>
      </country>
    </address>
  </contactInfo>
</organi sati onDetail s>
</provi der>
<versi onCreated>2010- 10- 19T10: 54: 04Z
</versi onCreated>
<fi rstCreated>2008- 11- 05T15: 22: 28Z</fi rstCreated>
<pubStatus qcode= "stat: usable" />
<servi ce qcode= "servi cecode: EUROVI SI ON">
  <name>Eurovi si on servi ces</name>
</servi ce>
<edNote>Original ly broadcast i n Germany</edNote>
<link rel= "itemrel: previ ew"
  href= "http: //www. exampl e. com/vi deo/2008- 12- 22/20081222- PNN- 1517-
  407624/i ndex. html "/>
</itemMeta>
<contentMeta>
  <contentCreated> 2008- 12- 22T23: 04: 00- 08: 00</contentCreated>
  <located type= "cptype: ci ty" qcode= "ci ty: 345678">
    <name>Berl in</name>
    <broader type= "cptype: statprov" qcode= "state: 2365">
      <name>Berl in</name>
    </broader>
    <broader type= "cptype: country" qcode= "country: DE">
      <name>Germany</name>
    </broader>
  </located>
  <creator qcode= "codesource: DEZDF">
    <name>Zwei tes Deutsches Fernsehen</name>
    <organi sati onDetail s>
      <locati on literal= "MAINZ" />
    </organi sati onDetail s>
  </creator>
  <contri butor qcode= "codeori gi n: DEZDF" role= "rol ecode: Techni cal Ori gi n">
    <name>Zwei tes Deutsches Fernsehen</name>
  </contri butor>
  <creator qcode= "codesource: GBRTV">
    <name>Reuters Televi si on Ltd</name>
  </creator>
  <language tag= "en" role= "langusecode: PartNarrated">
    <name>Engl ish</name>
  </language>
  <genre qcode= "genre: bi og">
    <name xml: lang= "en">Bi ography</name>
    <name xml: lang= "fr">bi ographi e</name>
  </genre>
  <subject type= "cpnat: abstract" qcode= "subj: 01000000">
    <name xml: lang= "en">Arts, Culture and Entertainment</name>
    <name xml: lang= "fr">Arts, culture, et spectacles</name>
    <narrower type= "cpnat: abstract" qcode= "subj: 01025000">
      <name xml: lang= "en">Ani mati on</name>
      <name xml: lang= "fr">Dessin ani mé</name>
    </narrower>
  </subject>
  <headline>Lori ot retrospecti ve</headline>
  <descripti on role= "descrole: dopesheet">
    Yesterday evening (November, 5) an exhibition opened in Berlin in
    honour of German humorist Vicco von Bülow, better known under the
    pseudonym "Lori ot", 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,
    televisi on sketches alongside late German actress Evelyn Hammann and
    a couple of movies. Under the name "Lori ot" in 1971 he created a
    cartoon dog named "Wum", which he voice acted himself. In 1976 the
    first episode of the TV series "Lori ot" was produced.
    <br />
  </descripti on>
  <descripti on role= "descrole: shotli st">
    Berl in, 21/12/2008
    <br />
  </descripti on>

```



```
- vs. Vicco von Bülow entering exhibition
<br />
- vs. Lorient and media
<br />
- 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 />
"Lorient is timeless. You always can watch him and I can always
laugh."
<br />
- actor Ulrich Matthes in exhibition
<br />
sot Ulrich Matthes, actor
<br />
" I would say: one of the great German classics. Goethe, Kleist,
Schiller, Thomas Mann, Lorient. That's the way I would say it."
<br />
</description>
</contentMeta>
<partMeta partid="Part1_ID" seq="1">
<icon href="http://www.example.com/video/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-000.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/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-447.jpeg"/>
<timeDelim start="447" end="831" timeunit="timeunit:editUnit"/>
<language tag="en" role="langusecode:VoiceOver" />
<description>Lorient and media </description>
</partMeta>
<partMeta partid="Part3_ID" seq="3">
<icon href="http://www.example.com/video/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-832.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/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-1082.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/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-1314.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/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-1617.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/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-2110.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/2008-12-22/20081222-PNN-1517-
407624/Keyframes/20081222-PNN-1517-407624-2733.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/2008-12-22/20081222-PNN-1517-
    407624/20081222-PNN-1517-407624.avi"
    format="fmt:avi"
    duration="111" durationunit="timeunit:seconds"
    videocodec="vcdc:c008"
    videoframerate="25"
    videoaspectratio="16:9" />
</contentSet>
</newsItem>

```

[<<return to Code Listings index](#)

9.3 Use case 2: Multiple renditions of web video and audio

Many applications require video and audio to be rendered and delivered in different formats for different play-out platforms. In this example, we show how NewsML-G2 can be used to provide a single piece of video content in various different formats, with each rendition giving technical metadata to assist the receiver in choosing the appropriate format(s).

This is achieved by repeating <remoteContent> wrappers. The examples show the following additional attributes from News Content Characteristics, which also describe the audio channel of each rendition:

9.3.1.1 Audio Bit Rate (@audiobitrate)

A positive integer indicating kilobits per second (Kbps)

9.3.1.2 Audio Sample Rate (@audiosamplerate)

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

9.3.1.3 Video Average Bit Rate (@videoavgbitrate)

A positive integer indicating the average bit rate in Kbps) of a video encoded with a variable fit rate.

LISTING 5 Multiple renditions of audio/video

```

<contentSet>
  <remoteContent
    href="http://www.example.com/video/2008-12-22/20081222-PNN-1517-
      407624/20081222-PNN-1517-407624-STREAM-700.FLV"
    size="8650645"
    contenttype="video/x-flv"
    width="240" height="180" duration="99"
    audiobitrate="64"
    audiosamplerate="44100"
    videoavgbitrate="700000"
    videoaspectratio="4.3" />
  <remoteContent
    href="http://www.example.com/video/2008-12-22/20081222-PNN-1517-
      407624/20081222-PNN-1517-407624-STREAM-80.3G2"
    size="1131023"
    contenttype="video/3gpp2"
    width="176" height="144" duration="99"
    audiobitrate="12"
    audiosamplerate="8000"
    videoavgbitrate="80000"
    videoaspectratio="16.9" />
</contentSet>

```

[<<return to Code Listings index](#)



This page intentionally blank

10 Packages

10.1 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 conveyed to the viewer in a named relationship, such as “Main Article”, “Sidebar”, “Background”. Another frequent application of packages is to aggregate content around themes, for example “Top Ten” news packages.

UK

BoE Rate Decision
in percent

Year	Rate (%)
2006	5.0
2007	5.5
2008	5.0
2009	1.5

Source: Bank of England

Bank cuts interest rates to record low
4:04pm GMT
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. [Full Article](#)

Government denies it will print more cash 11:17am GMT
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. [Full Article](#)

Rugby's Mike Tindall banned for drink-driving 2:02pm GMT
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. [Full Article](#)

MORE UK NEWS

- Crunch forces employees to work unpaid overtime
- Government warns of tax fraudsters
- Nissan to cut 1,200 jobs at Sunderland plant
- Sainsbury sales tops forecast
- More UK News...
- Cause of wind turbine damage unknown | Video
- Muslims warn Gaza crisis could provoke extremism
- Banks hiring young Britons to prepare for upturn

Figure 7: A “Top Ten” package on the Web

Source: Thomson Reuters

Packages can range from simple collections to rich hierarchical structures. Some sample applications of packages:

- ❖ A free mixture of different genres and types of media, grouped around a specific theme or subject. These may be created automatically by software acting on metadata, or by journalists exercising their professional judgement. For example, a package based on a major news event such as the inauguration of a world leader could include reporting of the event, with pictures, video, audio and graphics, together with background stories, biographies and pictures of the principal parties, references to previous similar events, and archive material of all types.
- ❖ A themed and ranked package of news stories and associated media, such as the top stories in the last hour; the top stories of the day. Additional filtering by human operator or software could create further packages such as the top business stories, top sports etc
- ❖ Packages with a specific application, such as a panel of EU related news for a web site which would have a pre-defined structure for the content. e.g. Top Story, Second and Third Stories, Top Video selection, Special Report, Fact of the Day.

NewsML-G2 is flexible in allowing a provider to package content that has already been published, or a package may be sent together with the content resources to which it refers (see [Exchanging News: News Messages](#)).

The IPTC recommends that providers use <packageItem> to convey formal relationships between News Items or other content resources, rather than <link>, which is designed to indicate supporting information or a resource that can be used by the human reader (e.g. as the original from which the current content was derived, or a link to a companion web site).

Links are not designed to be interpreted, but passed on with minimal processing. There is an interoperability issue: if a software vendor provides a NewsML-G2 processor, the processing rules should expect an arriving News Item to convey a single item of content; how could a processor know that the sender had used links with the intention of creating a pseudo-package? See [Packages and Links](#) for a detailed comparison of the <link> property.

Some characteristics of Packages:

- ❖ They always include content by reference – content is not conveyed inline in Package Items
- ❖ they promote content re-use – a single piece of content may be managed separately and referenced by multiple packages
- ❖ they can express structure, allowing news to be packaged as a named hierarchy of content resources
- ❖ packages can be managed (i.e., updated and versioned)
- ❖ references within a package may be ordered (or not), be complementary to its peers, or alternative.
- ❖ the relationships between content resources, as expressed by the package, can be given a period of validity.

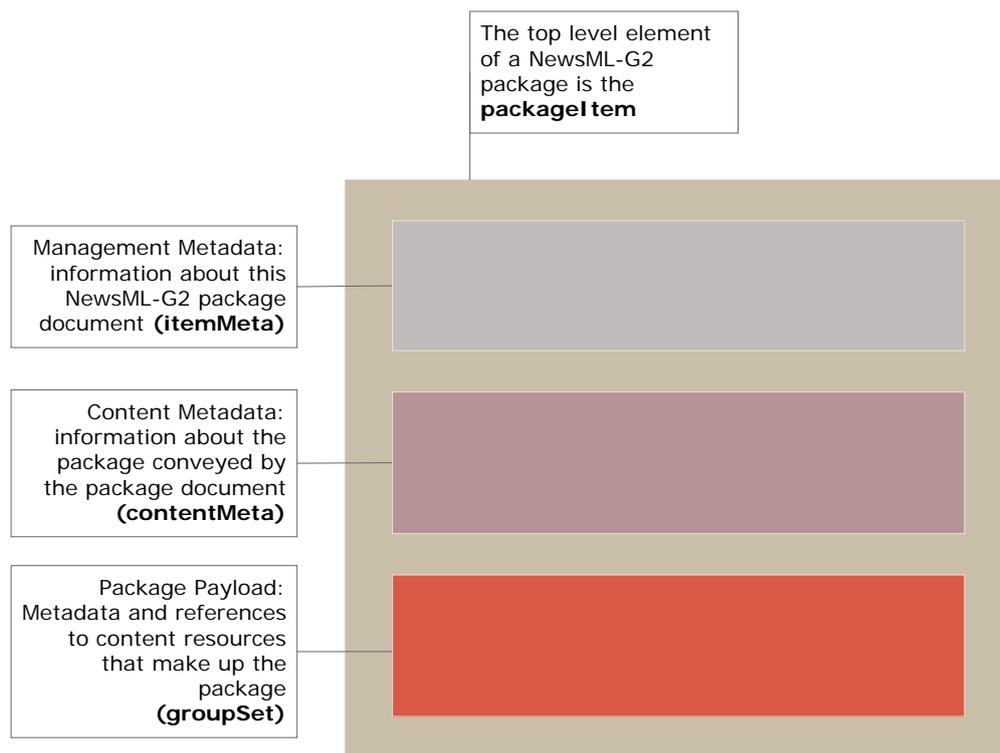


Figure 8; Structural model of a NewsML-G2 Package Item resembles the News Item



Please note that in general, the use of “item” with lower-case “i” is used generically to describe any discrete instance of content. A NewsML-G2 Item (News Item, Package Item etc.) is specifically indicated by capitalizing the word “Item”

Package Items share properties with News Items; compare the model as shown in [Figure 8](#) to the News Item model shown in [Figure 4](#) and they are seen to be similar.

Items and resources in Packages are ALWAYS included by reference – this enables the <packageItem> wrapper to be managed independently of the content. This does not in practice mean that the content must be physically separated from the package at the point of exchange; the <newsMessage> wrapper may be used to bundle package and referenced resources together in a single delivery. (See [Exchanging News: News Messages](#)).



The IPTC recommends that Package Items should reference NewsML-G2 Items if they are available (typically these would be 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.

10.2 Conveying Package Structures

As will be shown below, packages can have a variety of structures. In order to help receiving applications to process the information, we use additional properties to qualify the package structure. One such is the @role attribute of the <group> property, a basic building block of package content. We use this to denote the role that the component plays in the package, such as “main”, “sidebar”, “topstory”.

We may also use the <profile> property in <itemMeta> to name a pre-arranged template or transformation stylesheet used to generate the package, e.g. “text and picture”, “textpicture.xml”. <profile> is a “versioned internationalised string” datatype, which allows the named template or stylesheet to be versioned, consistent with denoting a software process:

```
<profile versionInfo="1.0.0.2">  
  text with one picture  
</profile>
```

Or

```
<profile versionInfo="1.0.0.2">  
  simple_text_with_picture.xml  
</profile>
```

See [Package Processing Considerations](#) for further discussion on this topic.

10.3 Simple Package Structure

The simplest possible Package would consist of one or more references to Items, using <itemRef> as a child element of <group>, as shown in [Figure 9](#) (a package consisting of one Item may be needed, in some circumstances).

```

<packageItem>
  <itemMeta.../>
  <contentMeta.../>
  <groupSet root="G1">
    <group id="G1" role="group:main">
      <itemRef..Item A ./>
      <itemRef..Item B ./>
      <groupRef idref="G2" />
    </group>
    <group id="G2" role="group:sidebar">
      <itemRef..Item C ./>
      <itemRef..Item D ./>
    </group>
  </groupSet>
</packageItem>

```

Figure 9: Simple Package using Item References <itemRef>

10.3.1 Group Set <groupSet>

The Item References wrapped by the Group element are in turn MUST be wrapped by <groupSet>. There MUST be only one Group Set per Package Item, and it MUST contain at least one Group. This structure allows Package Item to support a hierarchical group payload. The Group Set MUST identify the name of the Group that is the root (or primary) Group of the Set, using @root, and Groups MUST use @id in order that one of them is identifiable as "root". Each Group MUST also use @role to declare its role in the Group Set. This is a QName value. Thus

```

<groupSet root="G1">
  <group id="G1" role="group:main">

```

10.3.2 Item Reference <itemRef>

The <itemRef> element identifies an item or a Web resource using the LinkType template (CCL), or Link1Type (PCL), which means that it uses the **Target Resource Attributes** group.

At CCL, these attributes enable us to identify and locate the resource, using @href and/or @residref; version the item using @version, and give processing or usage hints using @contenttype, @format, and @size.

The referenced resources in the following package examples could have been simple resources such as text articles and images (see note above), but as recommended we will use <itemRef> to refer to News Items. As these are managed objects, we use @residref to identify and locate the referenced items.

In the examples, each referenced item has a Content Type of "application/vnd.iptc.g2.newsitem+xml", the registered IANA MIME type of a NewsML-G2 News Item. The following are the allowed values for @contenttype referencing Items:

Item	MIME type
News Item	application/vnd.iptc.g2.newsitem+xml
Package Item	application/vnd.iptc.g2.packageitem+xml
Concept Item	application/vnd.iptc.g2.conceptitem+xml
Knowledge Item	application/vnd.iptc.g2.knowledgeitem+xml
Planning Item	application/vnd.iptc.g2.planningitem+xml ¹¹

¹¹ Registration applied for and pending as of January 2011

@format may optionally be used to refine the MIME type, using a QCode. This may be used to give the processing application further information about the type of content identified by <itemRef>. In the case of a News Item, this could reflect the Item Class of the News Item, e.g. text, picture, video.

<itemRef> may have a @rel attribute to describe the relationship between the current Item and the target resource. It may use a <title> child element – the aim is to allow the Item Reference to inherit the Title of the target resource.

This relationship is expressed in NewsML-G2 using <groupSet>, <group> and <itemRef>, as shown in the code listing below.

LISTING 6 Simple Group Set at CCL

```
<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">
      <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="30039">
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
  </group>
</groupSet>
```

[<return to Code Listings index](#)

This relationship between the Group Set and the Group containing the Item References is illustrated as a diagram below.

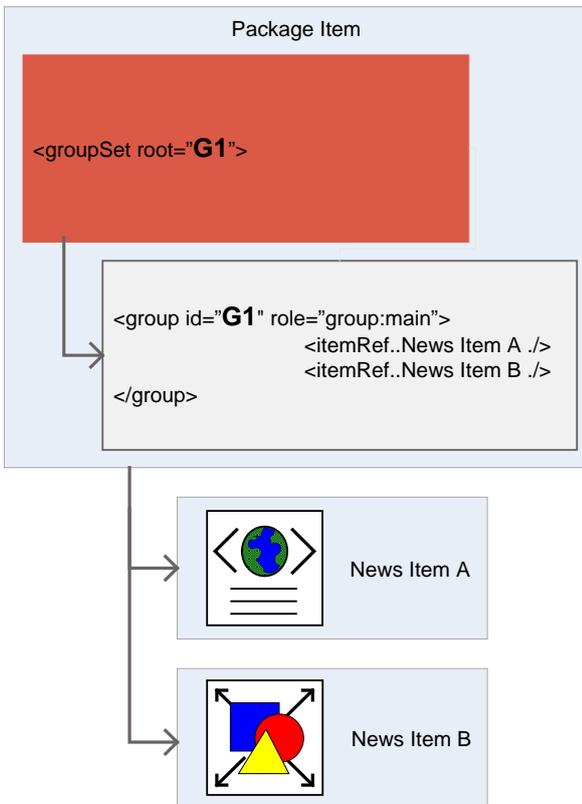


Figure 10: Simple Package Relationship

At PCL, the Link1Type data type's Hint and Extension Point allows us to use further properties of the target resource as child elements of the Item Reference. One use of this feature is to include processing hints for the receiving application, for example, by including the <itemClass> property, we can inform the receiving application of the type of content wrapped by the target resource

Inheriting an item's metadata may also allow the receiving application to use the headline and description (and other descriptive metadata) of the target resource, without the need to fully de-reference and process the resource.

Note that the additional properties do NOT have to be extracted from the target resource, but MUST comply with the structure of the target resource.

When using the Hint and Extension Point, the immediate child properties of <itemMeta>, <contentMeta> or <concept> can be used without the parent element; the additional properties are simply used as child elements of <itemRef>. For other wrapper elements, such as <partMeta> the full XML path, excluding the root (e.g. <newsItem>) element MUST be given. (See [Hint and Extension Point](#))

LISTING 7 Simple Group Set at PCL

```
<groupSet root="G1">
  <group id="G1" role="group:main">
    <itemRef resi:dref="urn:newsml:iptc.org:20081007:tutorial-item-A"
      contentType="application/vnd.iptc.g2.newsitem+xml"
      size="2345">
      <itemClass qcode="inat:text" />
      <provider literal="AFP"/>
      <pubStatus qcode="stat:usable"/>
      <title>Obama annonce son équipe</title>
      <description role="drol:summary"><p>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.
      </p>
    </description>
    </itemRef>
    <itemRef resi:dref="urn:newsml:iptc.org:20081007:tutorial-item-B"
      contentType="application/vnd.iptc.g2.newsitem+xml"
      size="300039">
      <itemClass qcode="inat:picture" />
      <provider literal="AFP"/>
      <pubStatus qcode="stat:usable"/>
      <title>Barack Obama arrive à Washington</title>
      <description role="drol:caption"><p>Si nous avons aujourd'hui un
        afro-américain et une femme dans la course à la présidence.
      </p>
    </description>
    </itemRef>
  </group>
</groupSet>
```

[<return to Code Listings index](#)



It would be possible to include <inlineXML> or <inlineData> as a child of <itemRef> and the document would validate against the NewsML-G2 Schema. However, this is strictly forbidden by the Standard as this would break the NewsML-G2 design and processing model. ONLY metadata properties, NOT content, may be inherited from the target resource.

10.3.3 References to Concepts <conceptRef>

Packages of news may also contain references to entities (people, organisations etc) and news events expressed as Concepts. For example, a provider could create a list of “see also” references to items, events and concepts that are relevant the main topic. The value to the end user of this “sidebar” is that it has been editorially selected; the references may not even have a relationship to the parent topic. that could be expressed using a “subject/related” reference within the parent item’s content metadata. Note also that this selection is transient – it only exists in the context of the package.

References to concepts are added to a Package using the <conceptRef> child element of <group> (See [14.5.3.1](#) for details of a Use Case for a Package of Event Concepts)

Note that individual Concepts can be exchanged either as part of a Knowledge Item or by a single Concept Item. It would be possible to use <itemRef> to (indirectly) reference a concept if it is contained in a Concept Item. However, throughout NewsML-G2, concepts are referenced using their Concept ID, a



QCode. It would be impractical to have to convert each concept to a Concept Item just to be able to reference it in a Package.

In the example below, there is a mix of Item References and Concept References in the same Package Group

LISTING 8 Simple Group Set including <conceptRef>

```
<groupSet root="G1">
  <group id="G1" role="group:main">
    <!-- MAIN GROUP -->
    <itemRef residref="urn. ...
  >
  </itemRef>
  ...
</group>
  <group id="G2" role="group:seealso">
    <!-- SIDEBAR GROUP -->
    <!-- AN ITEM (NEWSITEM) -->
    <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-A"
      contenttype="application/vnd.iptc.g2.newsitem+xml"
      size="2345">
      <title>Obama annonce son équipe</title>
    </itemRef>
    <!-- A CONCEPT (EVENT) -->
    <conceptRef qcode="iptcevents:20081007135637.12">
      <name>Barack Obama arrive à Washington</name>
    </conceptRef>
    <!-- A CONCEPT (PERSON) -->
    <conceptRef qcode="iptcpeople:20111007139337.15">
      <name>Hillary Clinton</name>
    </conceptRef>
  </group>
</groupSet>
```

[<<return to Code Listings index](#)



Concept References in a Package Item are transient and the relationship is implied but must be defined outside NewsML-G2. This is convenient and of value as a “one-time” selection by editors for human consumption. To express permanent and semantically-rich relationships between Items and Concepts, use properties such as subject and related.

10.3.4 Additional Properties of <group>

A rich set of additional metadata is available for <group>. Three global elements, i.e. properties used by other NewsML-G2 wrapper elements such as <contentMeta>, are re-used in a <group>:

Property Name	Definition
Title <title>	A short natural language name for the group
Signal <signal>	An instruction to the processor that the group requires special handling
Editorial Note <edNote>	Human-readable note about the group – not intended for publication

There is also an extension point to allow the inclusion of properties from non-IPTC namespaces.

10.4 Hierarchical Package Structure

Using multiple Groups we can create a hierarchy of Item references. For example, if we have a main text story with a picture, and a sidebar, or subsidiary text story with a picture, we can express this relationship which is shown in the diagrams below (**Figure 11** and **Figure 12**) and in code in **LISTING 9**.

In this example, we must associate the “sidebar” as a subsidiary of “main” by putting a reference to the “sidebar” Group with id=“G2” inside the “main” group so that it become a child of “main”. ALL groups

must be referenced somewhere in the Group Set; in the example below, there MUST be a reference to Group id="G2", otherwise it would be an "orphan" Group and a receiving processor would ignore it.

This relationship is expressed using <groupRef>, which enables Groups to be referenced from within other Groups using @idref, as illustrated by the following diagrams:

```

<packageItem>
  <itemMeta.../>
  <contentMeta.../>
  <groupSet root="G1">
    <group id="G1" role="group:main">
      <itemRef..Item A ./>
      <itemRef..Item B ./>
      <groupRef idref="G2" />
    </group>
    <group id="G2" role="group:sidebar">
      <itemRef..Item C ./>
      <itemRef..Item D ./>
    </group>
  </groupSet>
</packageItem>

```

Figure 11: Code outline of Hierarchical Package with two Groups

This creates a parent-child relationship between the two Groups of the Group Set as shown below

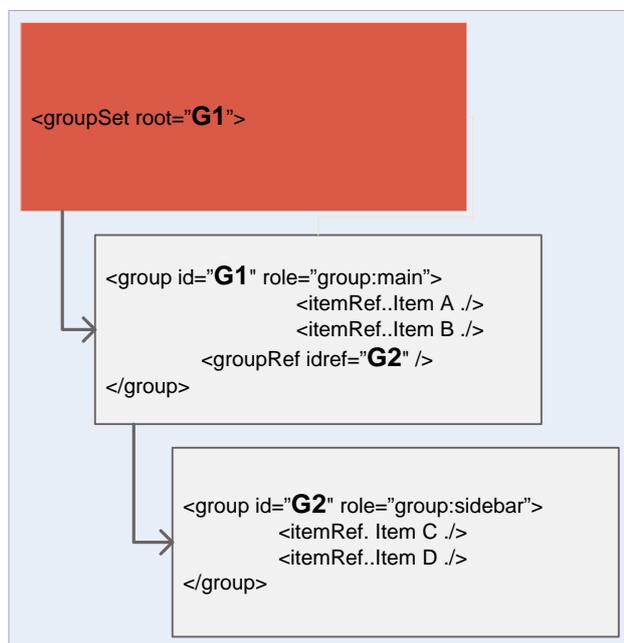


Figure 12: Relationship diagram of Hierarchical Package

The following code listing shows how the package contents and their relationships are expressed in NewsML-G2:

LISTING 9 Hierarchical Package

```

<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">
      <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="300039">
  <title>Barack Obama arrive à Washington</title>
</itemRef>
<groupRef idref="G2" />
</group>
<group id="G2" role="group:sidebar">
<itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-C"
  contenttype="application/vnd.iptc.g2.newsitem+xml"
  size="1503">
  <title>Clinton reprend son rôle de chef de la santé</title>
</itemRef>
<itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-D"
  contenttype="application/vnd.iptc.g2.newsitem+xml"
  size="350280">
  <title>Hillary Clinton à une rassemblement à New York</title>
</itemRef>
</group>
</groupSet>

```

[< return to Code Listings index](#)

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 story and picture of Hillary Clinton. It is referenced by a <groupRef> in Group G1.

10.5 List Type Package Structure (ordered, bag, alternative)

LISTING 9 can be re-cast from a hierarchical structure into a list, expressing the fact that the “main” group and the “sidebar” group are peers, rather than parent-child, using @mode to denote the relationship between them.

The Package Mode sets the context of the components of the group, which can be referenced Items, non NewsML-G2 resources, or other groups. The @mode attribute has one of three values, according to the IPTC Package Group Mode NewsCode:

- ❖ seq – denotes a sequential package group in descending order. Each component complements its peers in the group. An example use case would be a “Top Ten” list: each sub-group would provide references to a text article and a related picture.
- ❖ bag – an unordered collection of components. Each component complements its peers in the group. An example use case would be different components of a web news page with no special order, as in the example below.
- ❖ alt – an unordered collection. Each component is an alternative to its peers in the group. An example use case could be the same coverage of a news event, supplied in different languages.

(Note: these modes align with RDF Container elements.)

In the example below, we create a third group – in concept a “master” group – which contains <groupRef> elements referencing the peer groups in the package, with @mode set to indicate a bag, or unordered, package. This becomes the root group, given a role of “group:root” referenced by <groupSet>.

```

<groupSet root="root">
  <group id="root" role="group:root" mode="pgrmod:bag">
    ...

```

Note from the examples that the scope of @ref and @idref is purely local to the NewsML-G2 document; we can freely re-use these ids in other NewsML-G2 packages.

The relationship is shown in the diagrams ([Figure 13](#) and [Figure 14](#)).

```
<packageItem>  
  <itemMeta.../>  
  <contentMeta.../>  
  <groupSet root="root">  
    <group id="root" role="group:root" mode="pgrmod:bag">  
      <groupRef idref="G1" />  
      <groupRef idref="G2" />  
    </group>  
    <group id="G1" role="group:main">  
      <itemRef...Item A />  
      <itemRef...Item B />  
    </group>  
    <group id="G2" role="group:sidebar">  
      <itemRef...Item C />  
      <itemRef...Item D />  
    </group>  
  </groupSet>  
</packageItem>
```

Figure 13: Using <groupRef> to create an unordered “bag” of groups of Items

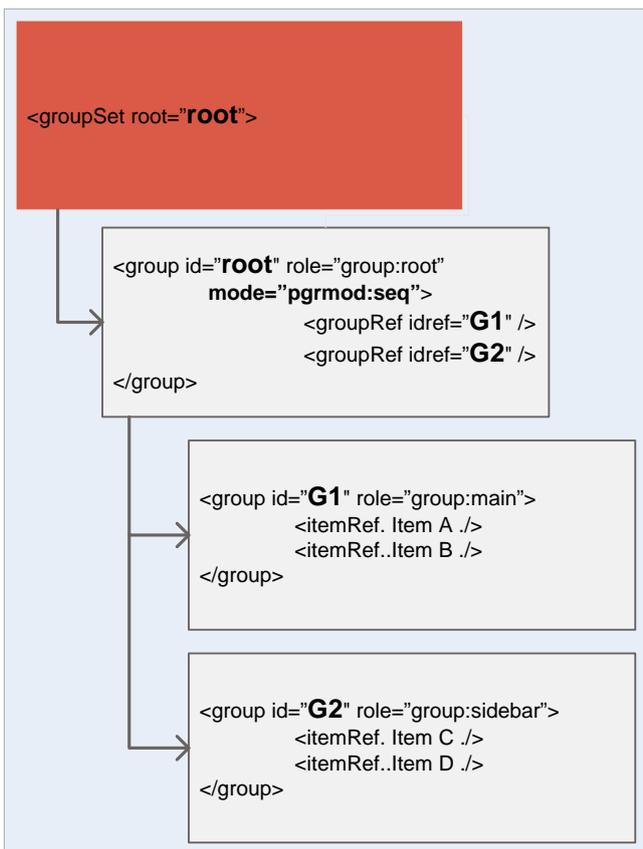


Figure 14: Groups G1 and G2 are both children of the Root Group



This relationship is expressed in NewsML-G2 as illustrated in the following code listing:

LISTING 10 List Type Package

```
<groupSet root="root">
  <group id="root" role="group:root" mode="pgrmod:bag">
    <groupRef idref="G1" />
    <groupRef idref="G2" />
  </group>
  <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">
      <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="300039">
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
  </group>
  <group id="G2" role="group:si debar">
    <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-C"
      contenttype="application/vnd.iptc.g2.newsitem+xml"
      size="1503">
      <title>Clinton reprend son rôle de chef de la santé</title>
    </itemRef>
    <itemRef residref="urn:newsml:iptc.org:20081007:tutorial-item-D"
      contenttype="application/vnd.iptc.g2.newsitem+xml"
      size="350280">
      <title>Hillary Clinton à une rassemblement à New York</title>
    </itemRef>
  </group>
</groupSet>
```

[< return to Code Listings index](#)

10.6 Package Processing Considerations

10.6.1 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 literal="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>
```

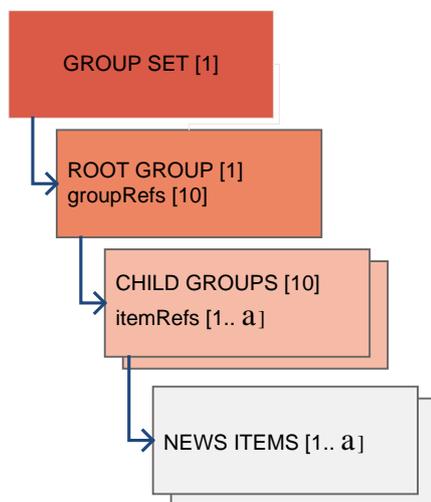
10.6.2 Facilitating the Exchange of Packages

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

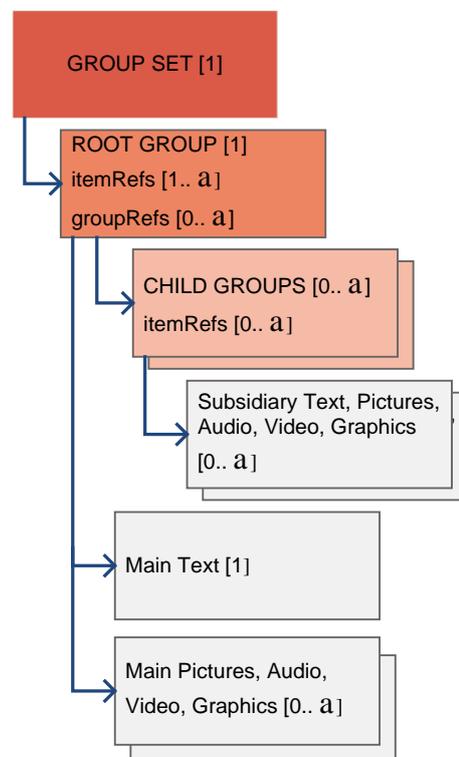


Figure 15: Diagrams of Package Profiles. The numbers in brackets indicate the required items

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 15**)

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"
/>
```

10.6.3 A "Top Ten" Package Example

In the example in **LISTING 11**, we will show an ordered list of references to News Items making up a Top Ten list of news stories on a themed topic, according to the Top Ten package profile shown on the left of **Figure 15**. This shows a "root" group that consists of a sequential list of child groups. Each child group contains a number of <groupRef> references to News Items, which convey different types of content (e.g. text, pictures).

The Package Item code in has a number of properties omitted for clarity and brevity, but some are worthy of note:

The @mode of the root group is set to "seq" as this is an ordered list of group references. Note that the @role of each sub-group is the SAME. This is because a group mode of "sequential" is logically a ranking of identical or similar types of component sub-groups. The IPTC recommends that the structures of the child sub-groups should be identical or highly similar when used in a sequential parent group.

The <service> property indicates the name of the service, using a QCode, to assist the receiver in routing the content to appropriate destination, and a natural-language <name> child element is also provided.

The name of the member of staff responsible for the package is given using a QCode to unambiguously identify him or her, together with their name and (at PCL) a @jobtitle QCode. We have also used <definition> and <note> as child elements to convey further information about the contributor that may be needed by the receiver.

As the person's role is time-bound, we qualify these two properties using @validto, which is set to the time the person leaves his or her duties for the evening, at which time the person filling the role of "Duty Packaging Editor" and their contact phone number may be expected to change, and therefore the current information about the Contributor would no longer be valid.

LISTING 11 The complete Top Ten style News Package

```
<?xml version="1.0" encoding="UTF-8"?>
<packageItem
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd
  guid="tag:example.com,2008:UK-NEWS-TOPTEN:UK20081220098658" version="17">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef
    href="http://www.example.com/customer/cv/catalog4customers-1.xml" />
  <itemMeta>
    <itemClass qcode="ni nat:composite" />
    <provider literal="MyNews" />
    <versionCreated> 2010-10-18T12:30:00Z</versionCreated>
    <firstCreated> 2008-12-20T12:25:35Z</firstCreated>
    <pubStatus qcode="stat:usable" />
    <generator versionInfo="3.0">MyNews Top Ten Packager</generator>
    <profile versionInfo="1.0.0.2">ranked_idref_list</profile>
    <service qcode="svc:uktopy">
      <name>Top Ten UK News stories hourly</name>
    </service>
    <title>UK-TOPTEN-NEWS</title>
    <edNote>Replace the previous version</edNote>
    <signal qcode="act:replacePrev" />
```



```

<link rel="i rel: previousVersion"
  residref="tag: example. com, 2008: UK- NEWS- TOPTEN: UK20081220098658"
  version="12" />
</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="root">
  <group id="root" role="group: root" mode="pgrmode: seq">
    <groupRef idref="G1" />
    <groupRef idref="G2" />
    <groupRef idref="G3" />
    <groupRef idref="G4" />
    <!-- group references G5 to G9 -->
    <groupRef idref="G10" />
  </group>
  <group id="G1" role="group: package">
    <itemRef residref="tag: example. com, 2008: TX- PAR: 20080529: JYC99"
      contenttype="application/vnd. iptc. g2. newsitem+xml" size="28047">
      <itemClass qcode="ni nat: text" />
      <provider literal="MyNews" />
      <versionCreated> 2008- 12- 20T16: 04: 00Z</versionCreated>
      <pubStatus qcode="stat: usable" />
      <headline>Bank cuts interest rates to record low</headline>
      <description role="drol: summary">
        LONDON (MyNews) - 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="tag: example. com, 2008: GX- PAR: 20080529: JYC44"
      contenttype="application/vnd. iptc. g2. newsitem+xml" size="23056">
      <itemClass qcode="ni nat: graphic" />
      <provider literal="MyNews" />
      <pubStatus qcode="stat: usable" />
      <headline />
      <description />
    </itemRef>
  </group>
  <group id="G2" role="group: package">
    <itemRef residref="tag: example. com, 2008: TX- PAR: 20080529: JYC80"
      contenttype="application/vnd. iptc. g2. newsitem+xml" size="2345">
      <itemClass qcode="ni nat: text" />
      <provider literal="MyNews" />
      <versionCreated> 2008- 12- 20T11: 17: 00Z</versionCreated>
      <pubStatus qcode="stat: usable" />
      <headline>Government denies it will print more cash</headline>
      <description role="drol: summary">
        LONDON (MyNews) - 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="tag: example. com, 2008: PX- PAR: 20080529: JYC34"
      contenttype="application/vnd. iptc. g2. newsitem+xml" size="24398">
      <itemClass qcode="ni nat: picture" />
      <provider literal="MyNews" />
      <pubStatus qcode="stat: usable" />
      <headline />
      <description />
    </itemRef>
  </group>
  <group id="G3" role="group: package">
    <itemRef residref="tag: example. com, 2008: TX- PAR: 20080529: JYC92"
      contenttype="application/vnd. iptc. g2. newsitem+xml" size="2345">
      <itemClass qcode="ni nat: text" />
      <provider literal="MyNews" />
      <versionCreated> 2008- 12- 20T14: 02: 00Z</versionCreated>

```



```

<pubStatus qcode="stat:usable" />
<headline>Rugby's Mike Tindall banned for drink-driving</headline>
<description role="drol:summary">
  LONDON (MyNews) - 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="tag:afp.com,2008:PX-PAR:20080529:JYC51"
  contenttype="application/vnd.iptc.g2.newsitem+xml" size="31285">
  <itemClass qcode="inat:picture" />
  <provider literal="MyNews" />
  <pubStatus qcode="stat:usable" />
  <headline>England international Mike Tindall</headline>
  <description role="drol:caption">
    England's Mike Tindall (L) scores his try despite a late
    tackle from Lionel Beauxis (R) of France during their Six Nations
    match at Twickenham March 11, 2007.
  </description>
</itemRef>
</group>
<group id="G4" role="group:package" />
<!-- more groups id="G5" to id="G9" -->
<group id="G10" role="group:package" />
</groupSet>
</packageItem>

```

[<<return to Code Listings index](#)

10.6.4 Packaging Other Resources

Although it is recommended that Package Items refer to NewsML-G2 Items, the @href property of <itemRef> allows implementers to refer to other types of resource, such as files on a local or remote file system. The example given in **LISTING 6** above is modified to use an @href to news objects as follows:

LISTING 12 Simple Group Set using @href at CCL

```

<groupSet root="G1">
  <group id="G1" role="group:main">
    <itemRef href="http://www.example.com/NewsML-G2/tutorial-item-A.xml"
      contenttype="application/ni tf+xml"
      format="fmt:ni tf"
      size="2345">
      <title>Obama annonce son équipe</title>
    </itemRef>
    <itemRef href="http://www.example.com/NewsML-G2/tutorial-item-B.jpg"
      contenttype="image/jpeg"
      format="fmt:jpeg"
      size="300039">
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
  </group>
</groupSet>

```

[<<return to Code Listings index](#)

Note that @contenttype is changed, and we have added @format as a further processing hint.



10.7 Packages and Links

Implementers may encounter the NewsML-G2 <link> property of Item Metadata and interpret it as a method of linking content in a relationship to create a “package” of news objects. The IPTC strongly recommends against this practice, for the reasons detailed below.

10.7.1 The difference in a nutshell

The IPTC’s intention in providing the <link> property that references supplementary resources, as well as the <packageltem> container that references Items (using <itemRef>) is:

- ❖ <link> is used to convey lightweight associated information about a NewsML-G2 Item
- ❖ A Package is a structured collection of references to resources, and the <packageltem> is a managed container used to deliver this collection as a Package; these features are not supported by links.

10.7.2 Purpose and features of the <packageltem>

News organisations and customers that create/consume news packages also have some or all of the following business requirements:

- ❖ Ability to manage the package itself, for example to identify and update it, and convey this information to partners.
- ❖ Need for the package to have metadata that is independent of its member objects, and which is stored and conveyed in a standard structure.
- ❖ Convey either a flat or a hierarchical structure of member item references and groups of member item references.
- ❖ Support different processing models, and convey “how to” processing information to the recipient.
- ❖ Independently publish/manage the package and the member items.

None of these requirements can be easily fulfilled using Links, which represent an unmanaged and unstructured “cloud” of objects related to a single Item.

For example, if a provider sends a text article with several pictures attached using links, an update which consists of a new picture would require a new version of the text article to be created and sent, even if the text remains untouched.

10.7.3 When to use <link>

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>.

10.7.4 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 [Hint and Extension Point](#).



10.7.4.1 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.

10.7.4.1.1 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)

10.7.4.2 Target Resource Attributes

See [8.3.3](#)

10.7.4.3 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>
```

10.7.4.4 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 [Supplementary information about a Concept](#)

10.7.5 Link examples

10.7.5.1 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 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" ← Item relation
      resIdref="tag:acmenews.com,2008:TX-PAR:20090529:JYC80"> ← Item ref
      <title>
        File picture of President Obama ← Title
      </title />
    </link />
  </itemMeta>
  ....
</newsItem>

```

10.7.5.2 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 guid="tag:acmenews.com,2008:TX-PAR:20090529:JYC85"
  version="1"
  standard="NewsML-G2" standardversion="2.9"
  conformance="power"
  ....>
  <itemMeta>
    <itemClass qcode="ni nat: text" />
    ....
    <edNote>With picture</edNote>
    <link
      rel="i rel: dependsOn" ← Item relation
      resIdref="tag:acmenews.com,2008:TX-PAR:20090529:JYC80"> ← Item ref
      <itemClass qcode="ni nat: picture" />
      <filename>
        obama-omaha-20090606.jpg ← File name
      </filename>
    </link>
  </itemMeta>
  ....
</newsItem>

```



```

</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>
      
      ...
    </inlineXML>
  </contentSet>
</newsItem>
  
```

10.7.5.3 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. The relationship between the current item and the previous version is "previousVersion".

```

<newsItem 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>



This page intentionally blank



11 Concepts and Concept Items

11.1 Introduction

Concepts in NewsML-G2 are a method of describing real-world entities, such as people 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 convey the meaning of news using concepts is a major driver of the migration 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 also used to convey event information, which is described in detail in [EventsML-G2](#).

11.2 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:

- ❖ "Jean-Claude Trichet" is a concept about which, or whom, we can express knowledge, for example, date of birth (December 20, 1942), 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: "Jean-Claude Trichet" *is a member of* "The European Central Bank" NewsML-G2 concept expressions thus conform with an RDF triple of subject, predicate and object

Concepts may be either "controlled" within the NewsML-G2 frame of reference, or they are "uncontrolled". Controlled concepts are managed by an "authority" (an organisation or company) and are maintained in Controlled Vocabularies that are managed using NewsML-G2. They are identified by a Concept URI, and their scope is global.

Uncontrolled concepts are identified by a literal string; their scope is local to the containing document. Although "uncontrolled" within the scope of NewsML-G2, they may be members of a Controlled Vocabulary that is managed outside the NewsML-G2 frame of reference, for example a provider could publish a list of codes in a printed document.

Every NewsML-G2 Concept identifier used must be unique in its scope, whether controlled or uncontrolled.

The Concept URIs identifying Controlled Concepts are expressed in NewsML-G2 using a special format: QCodes. A Concept URI must be a URL and it should resolve to natural-language and machine-readable information about the concept.

This Chapter will describe Controlled Concepts.

11.3 Concept Item

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. As shown in [Figure 16](#) it uses the same template as News Item ([Figure 4](#)) and Package Item ([Figure 8](#)) and therefore uses the same methods for identification, versioning, conformance levels, item metadata etc. Administrative metadata is supported for a Concept

Item, but not descriptive metadata, as the content of the concept speaks for itself and should be machine-readable.

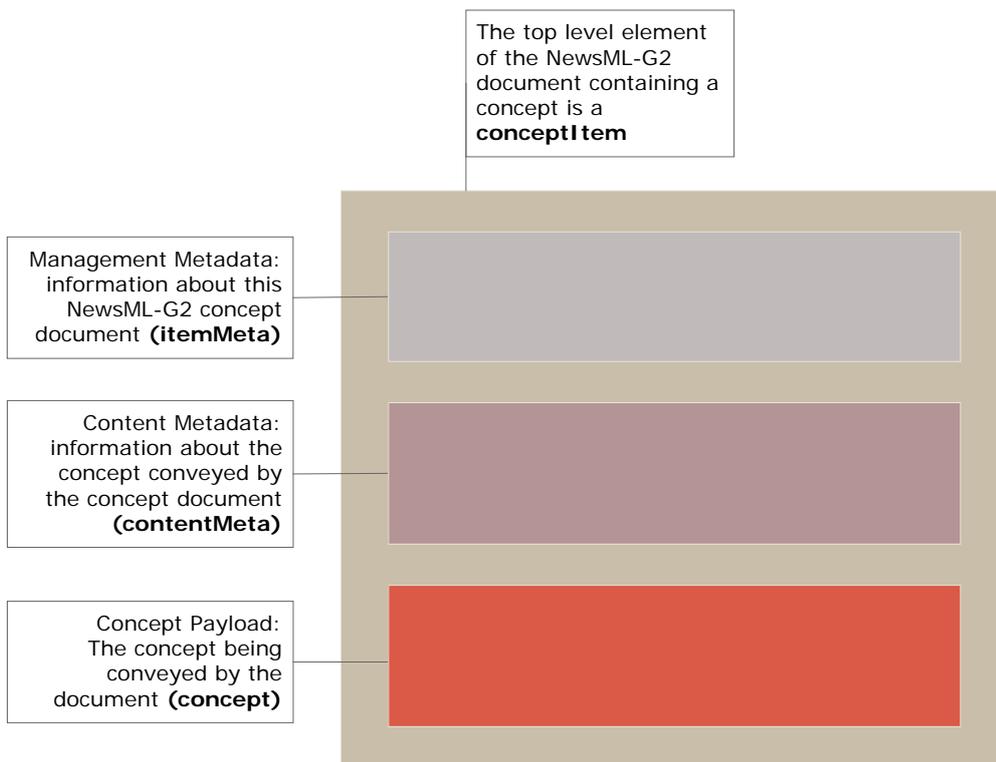


Figure 16: Concept Items share the same template as News Items and Package Items

11.4 Creating Concepts – the <concept> element

To create a concept for a real-world entity or some abstract notion, we need to define the concept using natural language descriptions and to identify it so that it can be used and re-used.

The <concept> element is a wrapper for the properties that express the concept in detail. The following elements are used to define a concept:

11.4.1 Concept ID <conceptId>

A concept MUST contain a <conceptId> which takes the form of a QCode. 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, the provider is indicating that it is no longer actively using this concept (it may have been merged with another concept; the knowledge base has been re-organised etc). Nevertheless, resources that were created before the change must be able to resolve the concept.

11.4.2 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>
    Jean-Claude Trichet
  </name>
</concept>
```



Concepts are designed to be useable in multiple languages:

```
<concept>
  <conceptId created="2000-10-30T12:00:00+00:00" qcode="subj: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>
```

11.4.3 Concept Type <type> and Related <related>

These optional elements express the “nature of the Concept” and additional characteristics of the concept. For example, we use the IPTC Concept Nature NewsCode to identify this concept is of type “person”. We can use Related to extend this notion into further characteristics of the concept.¹²

Both properties demonstrate the use of the subject, predicate, object triple of RDF to express a named relationship with another concept. The difference between the two properties in application is that <type> can only express one kind of relationship – “is a”. It is used to express the most obvious, or primary, inherent characteristic of a concept, as in:

“Jean-Claude Trichet” (Subject) “is a” (Predicate) “person” (Object).

This is represented using a QCode:

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

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).

11.4.3.1 Expressing relationships to other concepts using <related>

Inherent characteristics of a concept, e.g. that a person is male or female, can be described in terms of a named relationship with another concept. The <related> property uses either a @qcode or @literal to express this relationship. If the related concept is identified by a QCode; each provider should provide its own CV for this purpose (discussed in [Knowledge Items](#))

The nature of the relationship must be described using the @rel attribute. It uses a QCode, taken from a controlled vocabulary (CV), to identify the relationship.

For example, <related> can express “Jean-Claude Trichet” (Subject) “has a gender of” (Predicate) “male” (Object) using a QCode type value for @rel:

```
<related rel="relation:genderOf" qcode="gender:male" />
```

Or, using a @literal, the same relationship would be expressed:

```
<related rel="relation:genderOf" literal="M">
  <name xml:lang="en">Male</name>
</related>
```

¹² Prior to NewsML-G2 v2.7 and EventsML-G2 v1.6, the <facet> property was used to describe inherent characteristics of a Concept. This property is now DEPRECATED and should no longer be used. The reason for the change was that in prior versions, either <related> or <facet> could be used, and there was no clear rule for the different use cases. It was therefore decided to simplify the Standard by using <related> only.

Additional relationships can be included, such as: “Jean-Claude Trichet” (Subject) “has an occupation of” (Predicate) “public official” (Object):

```
<related rel="relation:occupationOf" qcode="jobtypes:puboff" />
```

11.4.3.2 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 using 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>
```

Note: The declaration of the XML Schema namespace in the above examples is purely for illustration.



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

The @value property is of type 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

11.4.4 Concept Definition <definition>

This is a block type element; it allows us to enter more extensive natural language information with some mark-up, if required. In this case, we can give a natural-language biography. Block type elements may use an optional @role QCode to disambiguate repeating Definition statements. We will use a QCode to indicate that this is the main definition. We could use additional definitions with @role values equivalent to, for example, “short definition”, “web definition” (a Block type element may contain hyperlinks). Note that the @role does NOT refer to the genre or nature of the contents of the Definition. A @role of “history” or “biography” i.e. describing the contents would be wrong.

```
<definition role="drol:main">
  Jean-Claude Trichet (born 20 December 1942) is a French civil servant who is the
  current president of the European Central Bank since 2003.<br />
  Trichet was born in Lyon, France and educated at the École des Mines de Nancy. He
```



```
later trained at the Institut d'etudes politiques de Paris and the Ecole nationale
d'administration, two higher education institutions with a tradition of providing
France's political and state administration elite.<br />
In 1993 Trichet was appointed governor of Banque de France and on November 1, 2003
he took Wim Duisenberg's place as president of the European Central Bank.<br />
</definition>
<definition role="drol:short">
Jean-Claude Trichet (born 20 December 1942, Lyon, France) is the
current president of the European Central Bank.
</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 this form.

11.4.5 Note <note>

Additional natural-language information on the concept may optionally added using this block type element, again with an optional @role:

```
<note role="nrol:disambiguation">
Not Jean-Claude Trichet the international powerboat racer
</note>
```

11.4.6 Remote Info <remoteInfo>

This is a Link Type (CCL) or Link 1 Type (PCL) wrapper has the same semantic as the <link> property (see 7.4.5 and 9.2.1.9), which is a child of <itemMeta>, i.e. to provide links to supplementary information.

For more information on the use of <remoteInfo>, see [Supplementary information about a Concept](#)

11.5 Relationships between Concepts

This is a group of four properties, <broader> <narrower> <related> and <sameAs> that enable the creation of particular types of relationship to similar or related concepts. For example, our subject was born in Lyon. We could create a concept for Lyon as follows, with a <broader> property that denotes that the city as part the department of Rhône:

```
<concept>
<conceptId qcode="urban:lyon" />
<type qcode="cpnat:geoArea" />
<definition role="drol:short">
Lyon, the second-largest French urban area, is the capital of the Rhône
department. It is a major industrial centre specialized in chemical,
pharmaceutical, and biotech industries. There is also a significant software
industry with a particular focus on video games.<br />
</definition>
<broader type="cpnat:geoArea" qcode="local:rhone">
<name xml:lang="fr">Rhône (département)</name>
</broader>
</concept>
```

<narrower> expresses the reverse relationship. A concept for Rhône could have a <narrower> property linking it to Lyon, and a <broader> 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 Jean-Claude Trichet that can be referenced using the appropriate alias.

```
<sameAs type="cpnat:person" qcode="AFPpers:567223">
<name>TRICHET, Jean-Claude</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="i so3166a2: DE" />  
<sameAs qcode="i so3166a3: DEU" />  
<sameAs qcode="i so3166n: 276" />
```

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

```
<related type="cpnat: person" qcode="people: 329465">  
  <name>Jean-Claude Trichet</name>  
</related>
```

The above example shows the use of <related> at CCL. Some providers wish to add further value to this concept relationship by indicating the nature of the relationship. At PCL, the property may be extended by adding @rel and/or @rank attributes. @rel is a QCode value that describes the relationship between the current concept and the target concept. @rank is a numeric ranking of the current concept amongst other concepts related to the target concept.

For example, using @rel we can denote that the European Central Bank "has a President" Jean-Claude Trichet. This relationship must be part of a CV of relationships, which might include "has a CEO", "has a Finance Director".

```
<related rel="relation: hasPresident" type="cpnat: person" qcode="people: 329465">  
  <name>Jean-Claude Trichet</name>  
</related>
```

If the European Central Bank is the second most important concept related to Jean-Claude Trichet, amongst other concepts related to him, we can denote this in the related property:

```
<related rank="2" rel="relation: hasPresident" type="cpnat: person"  
qcode="people: 329465">  
  <name>Jean-Claude Trichet</name>  
</related>
```

11.6 Entity Details

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.

11.6.1 Person Details <personDetails>

A concept of type "person" may hold the following additional properties:

11.6.1.1 Born <born> and Died <died>

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

```
<born>1942-12-20</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>1942</born>
```

The TruncatedDateTime property enables great flexibility in the expression of dates, especially where the precise date-time is either not known or not required. For example, some major historical events are often denoted as having occurred in a given year – the actual date is either not appropriate or not significant.

11.6.1.2 Affiliation <affiliation>

An affiliation of the person to an organisation.

```
<affiliation type="orgnat: bank" 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.



11.6.1.3 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”

11.6.1.4 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 either a QCode or Literal value
Area	<area>	Flexible Property	
Country	<country>	Flexible Property	
Postal Code	<postalCode>		

For example:

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



11.6.2 Organisation Details <organisationDetails>

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

11.6.2.1 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 [11.6.1.1](#))

11.6.2.2 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:headoff" literal="Frankfurt"/>
```

Or

```
<location type="loctypes:regoff" qcode="poi:75001">
  <name>Paris</name>
</location>
```

11.6.2.3 Contact Information <contactInfo>

Contact information associated with the organisation, uses the same structure as described in [11.6.1.4](#).

11.6.3 Geopolitical Area Details <geoAreaDetails>

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

11.6.3.1 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

11.6.3.2 Founded

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

```
<founded>1998</founded>
```



11.6.3.3 Dissolved

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

11.6.4 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 following additional properties:

11.6.4.1 Address

The location of the point of interest expressed as a postal address. The <address> element is a wrapper for child elements described in [11.6.1.4](#). 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 “how to get in contact” with the POI, so would be an office or mailing address.

11.6.4.2 Position <position>

The coordinates of the location as described in [11.6.3.1](#)

11.6.4.3 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>
```

11.6.4.4 Capacity <capacity>

The capacity of the POI is expressed as a Label:

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

11.6.4.5 Contact Information <contactInfo>

Contact information for the POI uses the <contactInfo> structure as described in [11.6.1.3](#).

11.6.4.6 Access details <access>

Methods of accessing the POI, including directions. This is a Block type of element, allowing some markup 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>
```

11.6.4.7 Detailed Information <details>

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

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

11.6.5 Object Details <objectDetails>

Examples of objects that may be expressed as a concept are works of art, books, inventions. The IPTC provides three properties 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:

11.6.5.1 Creation Date <created>

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

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

11.6.5.2 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.

11.6.5.3 Copyright Notice <copyrightNotice>

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

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

11.7 Code listings

The following listings show how Concept Items share the same generic template (anyItem) with other NewsML-G2 Items such as News Item and Package Item.

The first is a Concept Item that describes one of the IPTC Subject NewsCodes:

LISTING 13 "Abstract" Concept

```
<?xml version="1.0" encoding="UTF-8"?>  
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  guid="urn:newsml:iptc.org:20080229:ncdci-subjectcode"  
  version="101018123521"  
  standard="NewsML-G2"  
  standardversion="2.9"  
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/  
    XSD/NewsML-G2_2.9-spec-All-Core.xsd ">  
  <catalogRef  
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.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="cpnat:concept" />  
    <provider literal="IPTC" />  
    <versionCreated>2010-10-18T12:35:21+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 Subject NewsCodes</title>  
  </itemMeta>  
  <concept>  
    <conceptId created="2000-10-30T12:00:00+00:00" qcode="subj: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="it">nessuno</note>
</concept>
</conceptItem>
    
```

[<<return to Code Listings index](#)

The next Listing is a (fictional) Concept Item that a provider might send to customers for inclusion in a locally-hosted scheme of "people in the news" so that Items containing references to the Concept can be resolved:

LISTING 14 "Named Entity" Concept

```

<?xml version="1.0" encoding="UTF-8"?>
<conceptItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  guid="urn:newsml:iptc.org:20080229:ncdci-person"
  version="1010181123618"
  standard="NewsML-G2"
  standardversion="2.9"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.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" />
    <providerLiteral="IPTC" />
    <versionCreated>2010-10-18T12:38:18Z</versionCreated>
    <firstCreated>2008-12-29T11:00:00Z</firstCreated>
    <pubStatus qcode="stat:usable" />
    <title xml:lang="en">Concept Item describing Jean-Claude Trichet</title>
  </itemMeta>
  <concept>
    <conceptId created="2009-01-10T12:00:00Z" qcode="people:329465" />
    <type qcode="cpnat:person" />
    <name xml:lang="en-GB">Jean-Claude Trichet</name>
    <definition xml:lang="en-GB" role="drol:biog">
      Jean-Claude Trichet (born 20 December 1942) is a French civil servant who is
      the current president of the European Central Bank since 2003.<br />
      Trichet was born in Lyon, France and educated at the École des Mines de Nancy.
      He later trained at the Institut d'etudes politiques de Paris and the Ecole
      nationale d'administration, two higher education institutions with a
      tradition of providing France's political and state administrative
      elite.<br />
      In 1993 Trichet was appointed governor of Banque de France and on November 1,
      2003 he took Wim Duisenberg's place as president of the European Central
      Bank.<br />
    </definition>
    <note xml:lang="en-GB" role="nrol:disambiguation">
      Not Jean-Claude Trichet, the international powerboat racer
    </note>
    <related rel="relation:occupation" qcode="jobtypes:puboff" />
    <sameAs type="otherprov:AFP" qcode="pers:567223">
      <name>TRICHET, Jean-Claude</name>
    </sameAs>
    <personDetails>
      <born>1942-12-14</born>
      <affiliation type="orgrole: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">jc.trichet.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 literal="Frankfurt am Main" />
            <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 literal="Frankfurt am Main" />
            <country qcode="ISOCountryCode:de">
                <name xml:lang="en">Germany</name>
            </country>
            <postalCode>D-60066</postalCode>
        </address>
    </contactInfo>
</personDetails>
</concept>
</conceptItem>
    
```

[<<return to Code Listings index](#)

11.8 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). Using links to other resources we can also add articles, pictures and other objects to the Concept.

The use of <link> as a child of <itemMeta> in a Concept Item creates a problem when a number of Concept Items with Links are aggregated into a Knowledge Item: only the metadata inside the <concept> wrapper is carried across into the Knowledge Item; the Item Metadata, and with it the Link, is discarded.

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.

Implementers may wish to standardise on this implementation of asserting a link, as it creates a consistent "template" and processing model that can be used either for standalone Concept Items or for concepts conveyed in Knowledge Items. Therefore:

```

<conceptItem | knowledgeItem
  guid="tag:acmenews.com,2008:TX-PAR:20090529:JYC100"
  version="1"
  standard="NewsML-G2" standardversion="2.9">
  <itemMeta>
    .....
  </itemMeta>
  .....
  <concept>
    <conceptId created="2009-01-10T12:00:00Z" qcode="people:329465" />
    <type qcode="cpnat:person" />
    <name xml:lang="en-GB">Jean-Claude Trichet</name>
    <definition xml:lang="en-GB" role="drol:biog">
      Jean-Claude Trichet (born 20 December 1942) is a French civil servant who is
      the current president of the European Central Bank since 2003. <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 Jean-Claude Trichet
    </title>
  </remoteInfo>
  </concept>
</conceptItem>
    
```

← NO <link>

← "link" start

← Item ref



```

</title>
</remoteInfo> ← "link" end
<personDetails>
  <born>1942-12-14</born>
  . . . .
  <contactInfo role="contactrole:official">
    . . . .
  </contactInfo>
</personDetails>
</concept>
</conceptItem | knowledgeItem>

```

Using the rules given in [Hint and Extension Point](#) when adding properties of the target NewsML-G2 Item, the parent property must be included if it is not from either <contentMeta> or <itemMeta>:

```

<remoteInfo
  rel="irel:seeAlso"
  contentType="video/mpeg">
  resIdref="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
<description> ← no parent
  ECB official video of Jean-Claude Trichet working with senior
  colleagues at the Bank
</description>
<partMeta partId="part1" seq="1">
  <description>The first part shows...</description> ← includes
</partMeta>                                     parent
<partMeta partId="part2" seq="2">
  <description>The second part shows...</description>
</partMeta>
</remoteInfo>

```

11.9 Concepts and Controlled Values

11.9.1 Concept Identifier and Concept URI

The concept expressed in [LISTING 14](#) may be part of a taxonomy maintained by the provider to add usefulness and value to their news output. As shown in earlier examples, this information about the news takes the form of a <subject> property contained in the <contentMeta> of a News Item. The example

```

<subject type="cpnat:person" qcode="people:329465">
  <name xml:lang="en-GB">Jean-Claude Trichet</name>
</subject>

```

denotes a subject contained in the content as a concept of type "person" whose natural language name is "Jean-Claude Trichet".

Every NewsML-G2 Concept that is represented by a QCode¹³ must be a member of at least one scheme, or controlled vocabulary (CV). (Please see [Controlled Vocabularies and QCodes](#) on for a further detailed discussion on QCodes)

Controlled vocabularies may be created and maintained by a standards body, trade association, news agency or other organisation. This organisation defines which scheme (or schemes) that a concept belongs to.

The same concept can belong to more than one scheme. For example, the same organisation may have Arnold Schwarzenegger as a member of a CV of movie stars AND of a CV of politicians.

A concept identifier must be a URI: this "Concept-URI" must be a URL. It is formed of three parts:

- ❖ a URI which locates the scheme (the CV) as a Web resource - the Scheme-URI
- ❖ the code value string which is appended to this Scheme-URI to complete the Concept-URI

For example, the Scheme URI for the IPTC Concept Nature NewsCodes is:

```
http://cv.iptc.org/newscodes/cpnature/
```

Appending the string "person" to the Scheme URI produces the Concept URI for the "person" concept:

```
http://cv.iptc.org/newscodes/cpnature/person
```

¹³ Concepts which are local to the scope of a NewsML-G2 document may be represented by a literal value.

If this URL is entered into a Web browser, the page shown in **Figure 17** is returned:

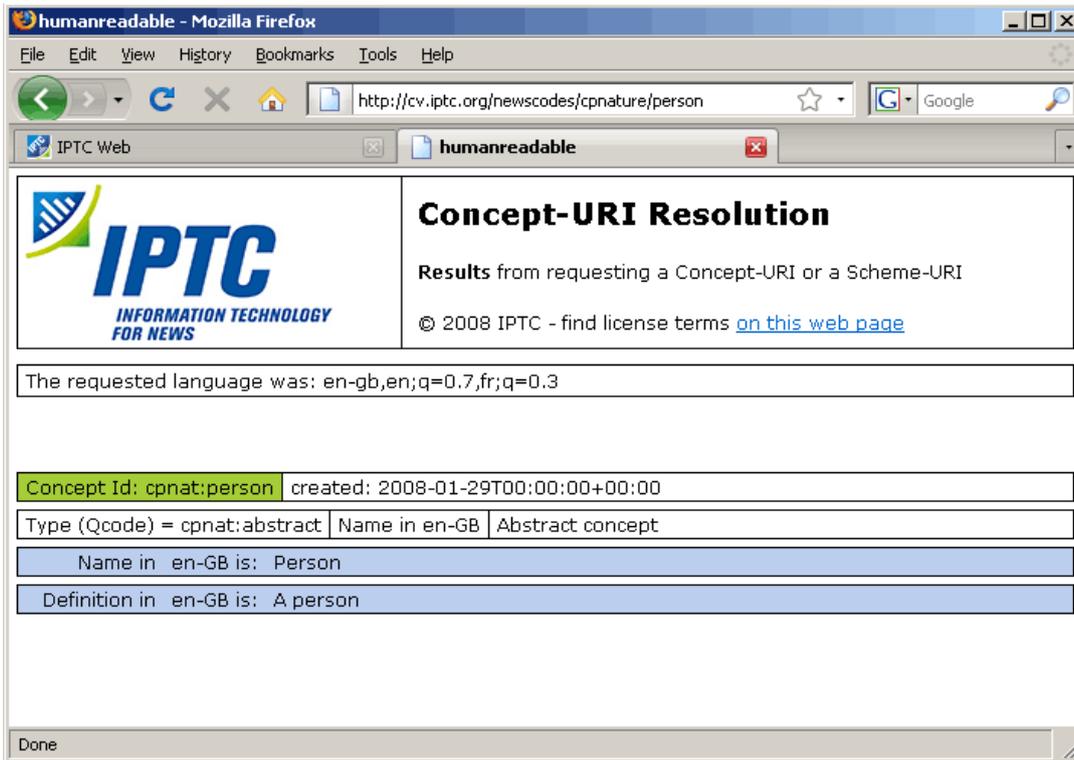


Figure 17: Entering the Concept URI in a browser returns human-readable information

11.9.2 From Concept URI to QCodes

A Scheme requires an abbreviation for this potentially long URI string otherwise the XML would become bloated and unwieldy. In the example of the IPTC Concept Nature NewsCode given above, the alias for the Scheme URI is “cpnat”. This is the “Scheme Alias”.

An alias for the Concept URI is formed by appending a colon (:) and an id to the Scheme Alias to produce, in G2 terminology, a “QCode”.

`qcode="cpnat: person"`

This QCode is an alias for the Concept URI that unambiguously identifies, within the scope of a document, the concept of “person” in the IPTC Concept Nature NewsCodes scheme.

11.9.3 Resolution of QCodes

Scheme URIs are contained in <catalog> statements in the root element of a NewsML-G2 document. A document MUST contain catalog information for the mandatory properties that require an IPTC NewsCode. A convenient way to do this is to use a remote catalog reference to the IPTC Catalog file:

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

where # is the version of the Catalog. From time to time, a new Catalog may be published, if a new scheme is added.

This file contains a number of Catalog statements:

```
<catalog xmlns="http://iptc.org/std/nar/2006-10-01/"
  additionalInfo="http://www.iptc.org/goto?G2cataloginfo"
>
  <scheme alias="acdc" uri="http://cv.iptc.org/newscodes/audiocodec/" />
  <scheme alias="prov" uri="http://cv.iptc.org/newscodes/provider/" />
  <scheme alias="scn" uri="http://cv.iptc.org/newscodes/scene/" />
  <scheme alias="subj" uri="http://cv.iptc.org/newscodes/subjectcode/" />
  .....
</catalog>
```



The alternative method of referencing the required NewsCodes would be to copy-and-paste the individual Catalog statements needed from the file hosted by the IPTC at the above URL and place them in a <catalog> in the document:

```
<catalog>
  <scheme alias="ni nat" uri="http://cv.iptc.org/newscodes/ni nature/" />
  <scheme alias="cpnat" uri="http://cv.iptc.org/newscodes/cpnature/" />
  <scheme alias="stat" uri="http://cv.iptc.org/newscodes/pubstat usg2/" />
</catalog>
```

Use the Catalog information, a processor can resolve the Scheme Alias part of a QCode, obtain the Concept URI, and thus unambiguously identify and potentially access the resource identified by the QCode.

11.10 Concepts in Practice

The more common method of exchanging Concept Items is as part of a Taxonomy (aka Controlled Vocabulary, 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

12 Knowledge Items

12.1 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 we have seen in the previous chapter **Concepts and Concept Items**, NewsML-G2 has powerful features for encapsulating this detailed information about entities and notions. Knowledge Items extend these features by allowing related concepts to be collated into a taxonomy.

For example, a profile of Russian Prime Minister 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.

Figure 18 shows a possible information flow for news information that exploits these possibilities.

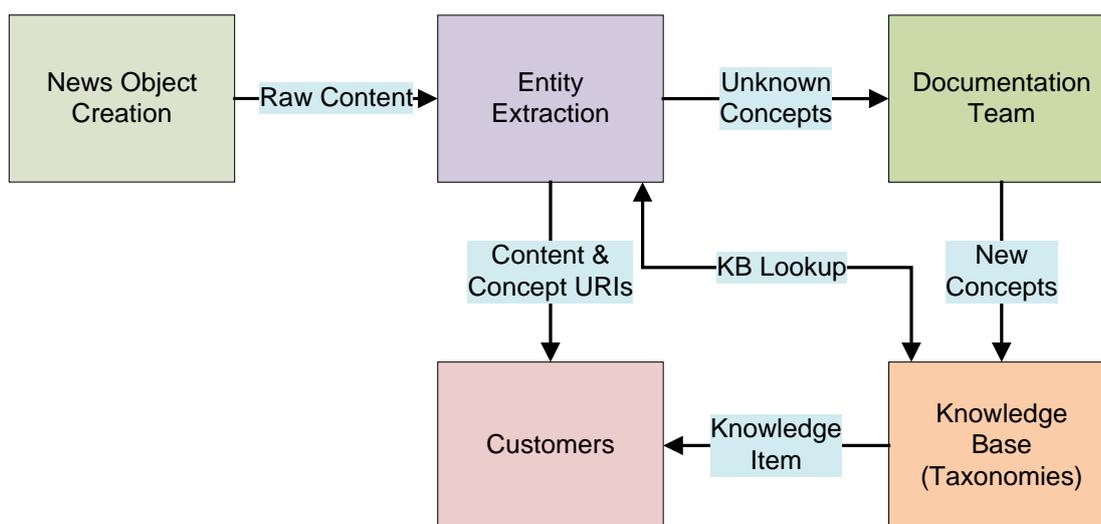


Figure 18: Information Flow for Concepts and Knowledge Items

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.

For example, IPTC NewsCodes are controlled vocabularies maintained as Knowledge Items. The contents of the CV are referenced using QCodes that uniquely identify concepts using a pair of values separated by a colon (:). The value to the left of the colon is the Scheme Alias. This is resolved using a Catalog



reference contained in the top-level element of the document which indexes the scheme alias to a URI: the Scheme URI. The Scheme URI must be a URL. The resource identified by the URL is a representation of this controlled vocabulary. It could be an HTML page for human readers, or a Knowledge Item as a machine-readable rendition. (see [Controlled Vocabularies and QCodes](#) for further discussion)

```
<itemClass qcode="ni nat: text" />
<catalog>
  scheme alias="ni nat" uri="cv/iptc.org/newscodes/ni nature/"
</catalog>
```

↑
Scheme URI found in the Catalog

The value to the right of the colon is an index into the scheme, which concatenated with the Scheme URI forms the Concept URI:

```
qcode="ni nat: text"
Scheme URI
+ "text"
= "cv/iptc.org/newscodes/ni nature/text"
```

↑
Concept URI

which resolves to one of a set of Concepts contained in the Knowledge Item:

```
<concept>
  <conceptId
    created="2008-01-29T00:00:00+00:00"
    qcode="ni nat: text" /> ← matching QCode
  <type qcode="cpnat: abstract" />
  <name xml:lang="en-GB">
    Text Item(s) ← Name
  </name>
  <definition xml:lang="en-GB">
    Text content in a News/PackageItem ← Definition
  </definition>
</concept>
```

Many more examples of Knowledge Items can be found on the IPTC web site at <http://www.iptc.org/newscodes/>. Choose View NewsCodes and follow instructions for downloading any of the CVs.

12.2 Structure and Properties

Knowledge Items share a common structure with News Items, Package Items and Concept Items. This is shown diagrammatically in [Figure 19](#), below.

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

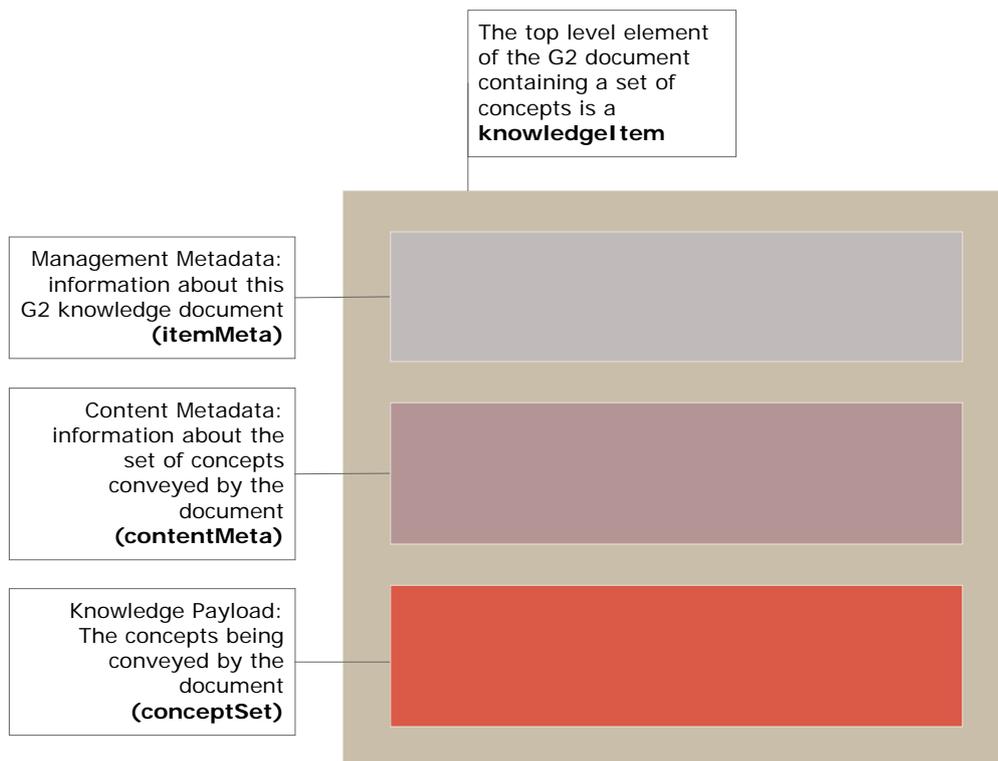


Figure 19: Structure of a Knowledge Item mirrors that of other NewsML-G2 Items

12.2.1 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"?>
<knowledgeItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd" ← XML Schema
  guid="urn:newsml:iptc.org:20090126:qqwpi ruuew4714" ← GUID
  version="1"
  standard="NewsML-G2"
  standardversion="2.2"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/
    catalog/catalog.IPTC-G2-Standards_16.xml" /> ← Remote Catalog
```

12.2.2 Management 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 a Knowledge Item.

```
<itemMeta>
  <itemClass qcode="cinat:scheme" />
  <provider literal="IPTC" />
  <versionCreated>2009-01-28T16:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
```

A more detailed discussion on the available properties of <itemMeta>, is contained in [Text](#).



12.2.3 Content Metadata

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

12.2.3.1 Administrative Metadata

Some date-time information may be needed:

```
<contentMeta>
  <contentCreated>2009-01-28T12:00:00Z</contentCreated>
  <contentModified>2009-01-28T13:00:00Z</contentModified>
  .....
```

12.2.3.2 Descriptive Metadata

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

<subject> One or more subject codes that are common to the Concepts conveyed in the Knowledge item. For example, a Knowledge Item conveying a set of concepts about France might have a subject code:

```
<subject type="cpnat:geoArea" qcode="ISO-3166-1-a2:fr">
  <name>France</name>
</subject>
```

<description> One or more free-form text descriptions of the collection of concepts in the Knowledge Item. The recommended IPTC Description Role NewsCode may be used:

```
....
<description xml:lang="en" role="drol:summary">
  A taxonomy of major concepts about the Republic of France
</description>
<description xml:lang="fr" role="drol:summary">
  Une taxonomie des principaux concepts de la République française
</description>
</contentMeta>
```

12.2.4 Concept Set

A single <conceptSet> element wraps zero or more <concept> components. The order of the Concepts is not important.

```
<conceptSet>
  <concept>
    <conceptId qcode="foo:bar1"/>
    <name>Concept One</name>
  </concept>
  <concept>
    <conceptId qcode="foo:bar2"/>
    <name>Concept Two</name>
  </concept>
</conceptSet>
```

Each <concept> component MUST have one <conceptId>, which must have a @qcode value denoting the "scheme alias:value" tuple that that will be used to identify the concept.

Each <concept> component MUST have a <name>. All other properties of a concept are optional and are fully described in **Concepts and Concept Items**.

The level of detail of information that a provider may make available in a Knowledge Item will depend on their 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.



12.3 Use Case: a Controlled Vocabulary 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, and since there is no IPTC NewsCode currently defined, any provider wishing to include this information would need to create their own Controlled Vocabulary (or perhaps share an agreed CV with other providers outside the auspices of the IPTC).

```
<accessStatus qcode="access: easy" />
```

The NewsML-G2 Specification recommends that when a receiving application processes a QCode, it SHOULD be able to resolve it to a resource (or resources) containing information about the concept that the QCode represents.¹⁴

If a customer receives event information containing the Access Status property and a QCode of "access:difficult", it is likely that the customer will want to know what this means in detail. It may be helpful to have this information available in several languages for an international product.

One way¹⁵ of doing this is would be to create and host a Knowledge item that would enable customer's receiving applications to resolve the QCode and take some action according to the information returned, for example to display an alert to the customer's event planning section.

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
	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	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.
	Accès Restreint	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	Access includes stairways with no lift or ramp available. It will not be

¹⁴ Processing rules and recommendations are detail in the NewsML-G2 Specification Document: Dealing with Controlled Values, available for download from the IPTC web site www.iptc.org

¹⁵ Not the only method – NewsML-G2 does not insist that such information must be permanently maintained in an XML file, only that an unambiguous identifier is provided for a scheme, which is made known to receivers.



Value	Names	Definitions
	difficult	possible to install bulky or heavy equipment that cannot be safely carried by one person
	L'accès est difficile	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

Each of these three values and definitions needs to be encapsulated as a Concept within the <concept> wrapper with a Concept ID and Name, for example

```
<concept>
  <conceptId qcode="access: easy" />
  <name xml:lang="en">Easy access</name>
```

We need to add the name in all of the chosen languages, identified by @xml:lang and the definitions:

```
<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>
```

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

The <knowledgeItem> wrapper, <itemMeta> and <contentMeta> are completed as documented above, with the following remarks:

- ❖ Since the conformance of the example is at the Core level (CCL), the @conformance attribute may be omitted from <knowledgeItem>
- ❖ The @xml:lang declaration in <knowledgeItem> is also superfluous in this instance,. There is no appropriate default language for the Knowledge Item.
- ❖ The <contentMeta> block contains a description in three languages of the purpose of the Controlled Vocabulary.

LISTING 15 Knowledge Item for Access Codes

```
<?xml version="1.0" encoding="UTF-8"?>
<knowledgeItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2_9-spec-A11-Core.xsd"
  guid="urn:newsml:iptc.org:20090202:ncdki-accesscode" version="20101018"
  standard="NewsML-G2"
  standardversion="2.9" >
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T00:00:00Z</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
      organisateurs.
    </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>

```



```
    </defini ti on>  
  </concept>  
</conceptSet>  
</knowl edgeI tem>
```

[<<return to Code Listings index](#)



13 Controlled Vocabularies and QCodes

13.1 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, 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 unique within the scope of the Item.

13.2 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 Subject NewsCodes 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.

13.3 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"
```

The generic term for a notion represented by a QCode is a *Concept*, (discussed in detail in [Concepts and Concept Items](#))

QCodes are used extensively in NewsML-G2 documents, but with an important distinction in their application, as follows:

- ❖ Many property attributes require a concept, represented by a QCode. For example the values of @role and @type attributes are represented by a QCode. These REFINE the value of the property.
- ❖ However, when a property has a @qcode, the value of the attribute IS the value of the property.

The Code Value suffix – in the example “usable” – is an index into the scheme of values represented by the Scheme Alias prefix (“stat” in this case).

13.3.1 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.

13.3.1.1 Catalogs

The <catalog> element may contain many <scheme> components, Catalog information is stored:

- ❖ 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_16.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.

13.3.2 Concept URI

Appending the QCode value suffix 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:

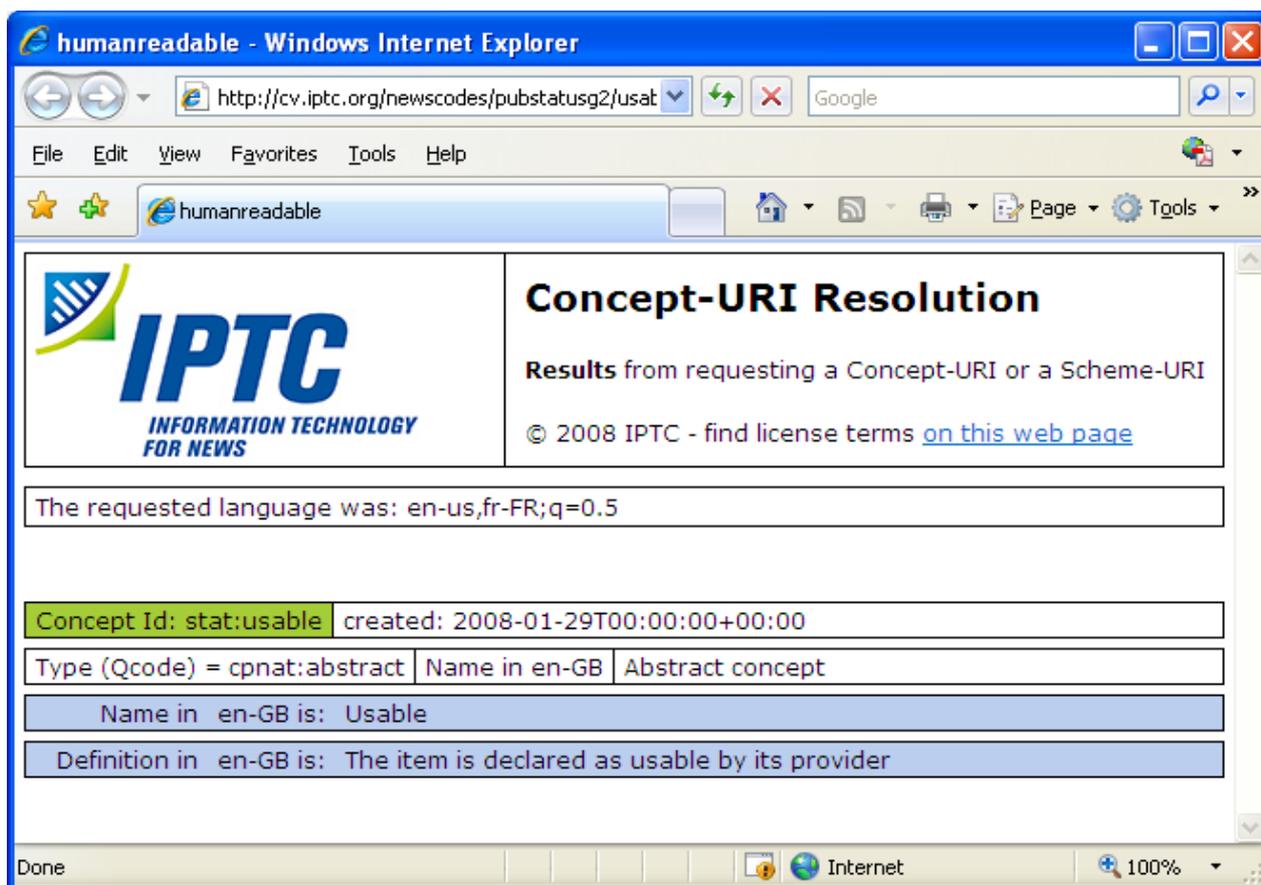


Figure 20: Human-readable browser page of a Concept URI

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. For example, 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.

However, the following example is WRONG 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:

```
<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 the above example, the processor can resolve both QCodes satisfactorily.

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

13.4 QCodes and Taxonomies

Also known as thesauri, knowledge bases and so on, taxonomies 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:rus12345">
```

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:rus12345”. Resolving this (fictional) scheme alias and suffix might result in the following concept URI:

```
http://www.example.com/knowledgebase/people/political_leaders/rus12345
```

and fetching the information at this resource:

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 Puti n
Summary	Became Russian prime minister in 2008 after serving two terms of office as President.
Background	Born 7 October 1952
Place of birth	Leningrad (St Petersburg)
Other	Speaks fluent German

The IPTC recommends that providers should make schemes containing concepts such as the above available to recipients as Knowledge Items.



13.5 Managing Controlled Vocabularies as NewsML-G2 Schemes

13.5.1 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 (13.5.2) and next creating a Knowledge Item from a set of existing Concept Items (13.5.3) 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 pair (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 a 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.

13.5.2 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/" />
  ...
```

If using a remote catalog, change the catalog URI to reflect a new version of the catalog (so that recipients know that they should update their cached 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 using Concept Items. 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"
  ...
```

This identifier resolves to the Concept URI `http://cv.example.org/schemeA/concept-x`

13.5.3 Creating a new Knowledge Item for distributing a CV

A Knowledge Item contains concepts from one to many Schemes. To create a KI:

- ❖ 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".
- ❖ Copy all or part of the selected concept details (the `<concept>` wrappers and their contents) into the Knowledge Item.

See [Chapter 12](#) for more information on the structure of Knowledge Items.



13.5.4 Maintaining a CV

13.5.4.1 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.

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 used for a different concept.
- ❖ Schemes themselves MUST NOT be deleted, because archived content is likely to use the concepts contained in a retired CV

13.5.4.2 Recommendations for non-complying Schemes

Below we provide guidelines for different Use Cases of a Scheme Authority failing to comply with the NewsML-G2 Specification where this is beyond the control of the end-user or the provider.

1. *The authority of the vocabulary governs the scheme URI and the code – but does not comply*
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 scheme URI of its own domain to make 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 to change the scheme URI in the case of the reuse of a code. This would avoid having the same Concept URI for different concept but would require a careful management of the vocabularies as actually a completely new Controlled Vocabulary is created by a new scheme URI.
3. *Another code is assigned to the same concept or a very likely 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 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>]

13.5.4.3 Changes to Catalogs

Catalog files only need to be changed when new Controlled Vocabularies are created and included in Items. Changes to the Scheme contents do not need to be reflected in a changed Catalog file, or <catalog> element within the Item.

Remote Catalogs **MUST** persist over time, so that if a scheme is deprecated, archived items that used that scheme will still be able to resolve scheme aliases and Concept URIs. Scheme Authorities **MUST** therefore issue an updated remote catalog under a new URL, and keep the URLs of previous remote catalogs indefinitely.

For example, the IPTC's catalogs have a URL of the form

"http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_#.xml" where # is an integer which is incremented by one each time an updated catalog is published.

Information about the latest version of the IPTC's Catalog file can be obtained from the IPTC web site www.iptc.org which also provides a link to download the latest file.



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.

13.6 Processing Controlled Vocabularies

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 they 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 file 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 URI.

13.6.1 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 News Item: text, picture, audio etc., using the News Item Nature NewsCodes.. (Concept Items use the complementary Concept Item Nature CV). 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 "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 be 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 Subject NewsCodes for classifying news content by subject matter, and that the scheme URI for these NewsCodes is "<http://cv.iptc.org/newscodes/subjectcode/>"
- ❖ The business requires that incoming content is routed to the appropriate department, according to the Subject NewsCodes found in the Items,
- ❖ A routing table is set up in the processor with a configurable rule "all items with a Subject NewsCode '04000000' to be routed to the department/workflow stage 'abc'."

How does the processor "know" that a <subject> property with a QCode containing "04000000" is an IPTC Subject NewsCode? The processor should not rely on the scheme alias "subj"; 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/subjectcode/" />
```

...this verifies that the QCode value is from the Subject NewsCodes scheme. A rule "all items with a Subject NewsCode of '04000000' to be routed to the Business News department" is satisfied and the NewsML-G2 Item processed appropriately.

13.6.2 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.

13.6.2.1 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.

13.6.2.2 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.

13.6.2.3 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.

13.6.2.4 Handling updates to Knowledge Items using <partMeta>

The question remains as to how the receiver can get information about *which* concepts have been modified (when cached concepts are synchronized with those in the latest received Knowledge Item).

This was addressed in NewsML-G2 v2.6 and EventsML-G2 v1.5 (onwards) by allowing <partMeta> to be used by any G2 Item, including Package Items and Knowledge Items.

The change also extended <partMeta> 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.

Taken together, these changes enable news providers to include update information about *individual concepts within the concept set* of a Knowledge Item. Before this change, it was only possible to provide update and version information about *the whole* Knowledge Item.

13.6.2.4.1 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 modification 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 [Updates Processing Notes](#))

13.6.2.4.2 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	Has modification timestamp	Compare modification timestamp:



Concept in local cache	Concept just received in KI	Processor should:
timestamp		i/ if the modification timestamp from the KI is <i>later</i> than the one from the cache: update the concept in the cache. ii/ if the modification timestamp from the KI is <i>earlier</i> 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. The concepts are identified using @id. <partMeta> elements refer to these @ids using the @contentrefs attribute and contain a <contentModified> value informing receivers of the timestamp (date-time) of the change.

As the receiver previously processed changes related to the modification in November 2009, it may now conclude that the concepts with the ids c01 and c02 have been modified in January 2010 and thus need to be updated in the cache.

```
<?xml version="1.0" encoding="UTF-8"?>
<knowledgItem ... >
  ...
  <itemMeta>
    ...
  </itemMeta>
  <contentMeta>
    ...
  </contentMeta>
  <partMeta contentrefs="c01 c02">
    <contentModified>2010-01-28T13:00:00Z</contentModified>
  </partMeta>
  <partMeta contentrefs="c03">
    <contentModified>2009-11-23T13:00:00Z</contentModified>
  </partMeta>
  <conceptSet>
    <concept id="c01">
      <conceptId qcode="access: easy" />
      ...
    </concept>
    <concept id="c02">
      <conceptId qcode="access: difficult" />
      ...
    </concept>
    <concept id="c03">
      <conceptId qcode="access: restricted" />
      ...
    </concept>
  </conceptSet>
</knowledgItem>
```

Developer note

[NB: The normalize-space trick works for XPath 1.0, but fails for XPath 2.0 – using XML Spy 2006]

If partMeta contains the following list of idrefs:

```
<partMeta contentrefs="c2 c1" />
<partMeta contentrefs="c11 c111 c1111" />
```

Then the following XPath will match both of the above:

```
partMeta[contains(@contentrefs, 'c1')]
```

To select the single partMeta which references 'c1', the following XPath idiom is suggested:

```
partMeta[contains(concat(' ', normalize-space(@contentrefs), ' '), 'c1 ')]
```

13.6.2.4.3 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.

13.7 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.¹⁶
- ❖ **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.

13.7.1 Solution for Case 1

The national news agency can create its own CV, containing all of the codes it wishes to use from the IPTC Scheme, and add its own names and definitions for the codes.

Then for each Item that uses the provider's CV, a child <sameAs> property is added to the <scheme> element in the <catalog>:

```
<catalog>
  <scheme alias="itemrel" uri="http://cv.example.org/codes/itemrel/" ← private CV
    <sameAs>http://cv.iptc.org/newscodes/itemrelation/</sameAs> ← original CV
  </scheme>
</catalog>
```

This <sameAs> asserts that the ALL of the codes in the private scheme are semantically the "same as" the corresponding codes in the original scheme identified by the Scheme URI.

The <sameAs> child element of a Scheme complements the <sameAs> child element of a Concept, and enables a provider to apply a Same As relationship at the level of a **set** of Concepts. This speeds up QCode processing: a processor does not have to check each individual concept for this Same As relationship but can apply this relationship directly by matching the scheme identifier of the concept (used as a property value) with the <sameAs> scheme. (Note, however, that this should never stop a processor from checking for other Same As relationships. The scheme <sameAs> is only one of potentially many Same As relationships which could be found at the individual Concept level.)

13.7.1.1 Scheme <sameAs> example

The provider sends a News Item that contains the following property in <itemMeta>:

```
<link>
  rel="itemrel:seeAlso"
  contentType="image/jpeg">
  residref="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
</link>
```

The processor resolves the QCode "itemrel:seeAlso" via the <catalog> statement which matches the scheme alias "itemrel" to the scheme URI "http://cv.example.org/codes/itemrel/", leading to the unique Concept URI of "http://cv.example.org/codes/itemrel/seeAlso".

This Concept URI resolves to the provider's private scheme which contains the following information (a version of the IPTC's Item Relation code "seeAlso" with name and definition in Czech):

```
<concept>
  <conceptId created="2010-01-29T00:00:00+00:00" qcode="itemrel:seeAlso" />
```

¹⁶ The IPTC hosts CVs (the NewsCodes) which contain concept information in several languages, such as English, French, Spanish and German. However, it is not practical – the IPTC does not have the resources – to provide concept details in every language being used by all news providers.



```
<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 processor is able to use the <sameAs> property of the News Item <catalog> to match this concept (itemrel:seeAlso -> http://cv.example.org/codes/itemrel/seeAlso) to the corresponding concept (irel:seeAlso -> http://cv.iptc.org/newscodes/itemrelation/seeAlso) in the original IPTC scheme:

```
<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 <br />
  </definition>
  ...
</concept>
```

13.7.1.2 Rules for Scheme <sameAs>

The semantics of <sameAs> are:

"all of the concepts in the scheme identified by the parent scheme/@uri have a 'same as' relationship to concepts with the same code in the original scheme identified by this <sameAs> statement."

In practice, this means:

- ❖ The Scheme identified by @uri (the *provider's* private scheme) must NOT use a code that does not exist in ALL of the *original* Scheme identified by the <sameAs> elements. In other words, a provider cannot add new concepts to its private Scheme that have the effect of extending 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).

Scheme <sameAs> 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.

13.7.2 Solution for Case 2

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 <broader> 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 Broader relationship to another concept is a new concept with semantics narrower than those of the broader concept.

13.7.2.1 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/wfroles/" />
</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:finalreview" />
    ...
  </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>
```



13.8 QCodes and non-URI characters

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.

When converting a pre-existing controlled vocabulary to NewsML-G2, some of the codes used in the CV may contain characters that lie outside the set of characters allowed in a URL, such as spaces. Or the CV may contain characters that are allowed but which are part of the set of reserved characters, such as #, %.

These “non-URI” characters are permissible as code values in QCodes, but must be escaped by the provider as per RFC 3986 using a percent-encoding scheme if resolving the scheme-code pair to a full Concept URI

The unreserved characters that are allowed 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 should be percent-encoded if they are being used for a purpose other than that intended in RFC3968. The reserved characters are:

! * ' () ; : @ & = + \$, / ? % # []

For example, a provider’s scheme has a concept for “three-month moving average of the stock exchange daily closing index” which is represented by legacy code #3FTSE

The catalog entry for the parent Scheme is (say):

```
<scheme alias="fc" uri="http://cv.example.org/schemes/fc/" />
```

The QCode for this concept consists of the Scheme Alias/Code pair separated by a colon. However, a code value of #3FTSE would resolve to a Concept URI:

http://cv.example.org/schemes/fc/#3FTSE

... which is not a valid URL.

The reserved character # needs to be percent-encoded (%23) thus:

http://cv.example.org/schemes/fc/%233FTSE

Note that the QCode for this concept in the G2 Item would be:

```
qcode="fc: #3FTSE"
```

as the percent-encoding is only performed when resolving the code to a Concept URI

13.8.1 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”.

13.8.2 White Space in Codes

Many codes in existing controlled vocabularies 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}) - must be 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.

13.8.2.1 Example use case

A provider delivered news using NewsML 1.2 which has no restriction on the use of space in codes. The provider now delivers news using NewsML-G2. How does a receiver match codes of the same CV delivered using NewsML 1.x against QCodes from the NewsML-G2 feed?

Example Code from existing CV: "ab 3x"

Provider uses the same CV but encoded for NewsML-G2 using "~" to replace spaces: "ab~3x"

Receiver decodes code for matching against the existing CV: "ab 3x"

13.9 Syntactic Processing of QCodes

This section provides a summary of the processing model. Please also read the NewsML-G2 Specification (download from www.iptc.org) for a full technical description.

13.9.1 Creating QCodes from Scheme Aliases and Codes

13.9.1.1 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})

13.9.1.2 Codes

The IPTC recognises the problem of legacy codes containing characters that require encoding before use within a URI, and has created an explicit processing model for Schemes and QCodes:

To create a QCode for an Item, use the following steps:

1. Concatenate the Scheme Alias, a colon and the Code to a string.
e.g. fôô:bår
2. Apply the resulting string as a QCode value.
e.g. <subject qcode="fôô:bår" />

13.9.2 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. Percent-encode the Code Value: bår → b%C3%A5r
6. Append the percent-encoded value to the Scheme URI to make the full Concept URI:
example: http://cv.example.org/cv/somecodes/b%C3%A5r
7. It is highly recommended to use only full Concept URIs to compare identifiers of concepts.



13.10 QCode and Literal Identifiers

The NewsML-G2 Standard recognises that it is not always possible to use QCodes, which are values from controlled vocabularies that can be resolved by a processor, as identifiers. 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.

For this reason, the NewsML-G2 designers conceived Flexible Property datatypes, that may have EITHER a @qcode or a @literal value, or none.

13.10.1 The difference in a nutshell

- ❖ A @qcode is a globally unambiguous identifier which via the scheme-code pair, resolves to a globally unique Concept URI that can be shared among all Items.
- ❖ A @literal is an identifier that may be from a CV managed outside NewsML-G2. In some circumstances it may be an unambiguous identifier within the scope of the containing Item that is not globally unique and is not shared with other Items. See [13.10.4](#)

13.10.2 What @literal is; what it is NOT

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, providers SHOULD add the <name> property. For example:

```
<subject type="cpnat:poi" literal="eiffeltower">
  <name xml:lang="en">The Eiffel Tower</name>
  <name xml:lang="fr">La Tour Eiffel</name>
</subject>
```

This is the IPTC's recommendation for inter-operability and language-independent processing, but is not mandatory. If no human-readable property is available, receivers MAY use the @literal value for display purposes as a last resort.

13.10.3 Use of @literal in a NewsML-G2 Item

A @literal value can be used to identify a concept within the scope of an Item. As a general rule identical values of @literal, when applied to different metadata properties, do NOT indicate that the identified concepts are identical. A receiver CANNOT infer anything about the relationship of concepts that have identical @literal values.

For example, the following is correct by NewsML-G2 rules, although the same @literal value "int001" obviously identifies two different concepts within the same Item:

```
<contentMeta>
  ...
  <located type="cpnat:poi" literal="int001">
    <name>Paris Hilton</name>
  </located>
  ...
  <subject type="cpnat:person" literal="int001">
    <name>Paris Hilton</name>
  </subject>
  ...
</contentMeta>
```

The first exception to this general rule is that if an <assert> element of a NewsML-G2 Item identifies a concept using a @literal value, as in the following example:

```
<contentMeta>
  ...
  <located type="cpnat:poi" literal="int001" />
  ...
  <subject type="cpnat:poi" literal="int001" />
  ...
</contentMeta>
<assert literal="int001">
  <name xml:lang="en">The Eiffel Tower</name>
```



```
<name xml:lang="fr">La Tour Eiffel</name>  
</assert>
```

In this case, the @literal value "int001" of the <located> and the <subject> properties MUST identify the same concept for which details are asserted by the <assert> element with the same @literal "int001",

When a @literal value is used as shown in the example above, without an <assert>, the receiver is not able to infer anything about the values of @literal: they may identify the same concept, or they may not.

The second exception to the general rule is that there could be a contract outside NewsML-G2 rules governing @literal values of specific metadata properties, like provider XY has a list of literal values which will be used with the <located> properties of Items released by this provider. .

So in the first code example above, a provider could, outside NewsML-G2, maintain a list of identifiers to be used with the <located> element, in which the value "int001" maps to the hotel, and another list of identifiers to be used with the <subject> element, in which "int001" maps to the person.

The rules for the use of @literal are set out in detail below.

13.10.4 Changes to @literal

In earlier versions (prior to NewsML-G2 v2.7 and EventsML-G2 v1.6) some implementers commented that the use of @literal identifiers in some properties was too strict and did not take account of some use cases. The issues to be addressed were:

- ❖ literal values could be defined in a provider's controlled vocabulary which is defined by other means than NewsML-G2
- ❖ In the absence of @qcode or <bag > a @literal was mandatory, but in some use cases, it is inappropriate and does not add value.

Therefore some statements about the use of @literal were redefined:

- ❖ 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:

1. 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.
2. When importing metadata the values of literals may contain codes which have not yet been checked to be from an identified vocabulary.
3. 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.



This page intentionally blank



14 EventsML-G2

14.1 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.

14.1.1 Business Advantages

EventsML-G2 shares a common structure with NewsML-G2¹⁷ 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.

¹⁷ 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 iCalendar¹⁸ 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.

14.1.2 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 G2 object.



As there is no longer a separate EventsML-G2 standard, and news event management is now also an integral part of NewsML-G2, we refer generically to G2 in this chapter.

There are two methods for expressing events using G2, each suited to a particular type of information, or application:

- ❖ Persistent event information that will be referenced by other events and by other G2 Items is instantiated as Concept Items (**Figure 21**), or as members of a set of event Concepts in a Knowledge Item.
- ❖ Transient event information that is “standalone” and volatile can be conveyed in NewsML-G2 using an <events> structure inside a News Item. (**Figure 22**)

Most of the properties of events are common to the two models, however, note the following key difference:

- ❖ Persistent Events are expressed as a Concept with a Concept ID, which may be persistently and unambiguously referenced by other Items over time, while
- ❖ Transient Events are expressed as an <event> in the <events> of a News Item, have no Concept ID and can therefore never be referenced from other Items.

¹⁸ A mapping of iCalendar properties to G2 properties can be found in the G2 Specification document on the IPTC web site (www.iptc.org) See also [IETF iCalendar Specification RFC 5545](http://www.ietf.org/rfc/rfc5545)

The use-case for managed (persistent) events would be where a provider makes each of announced events uniquely identifiable by receivers. 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.

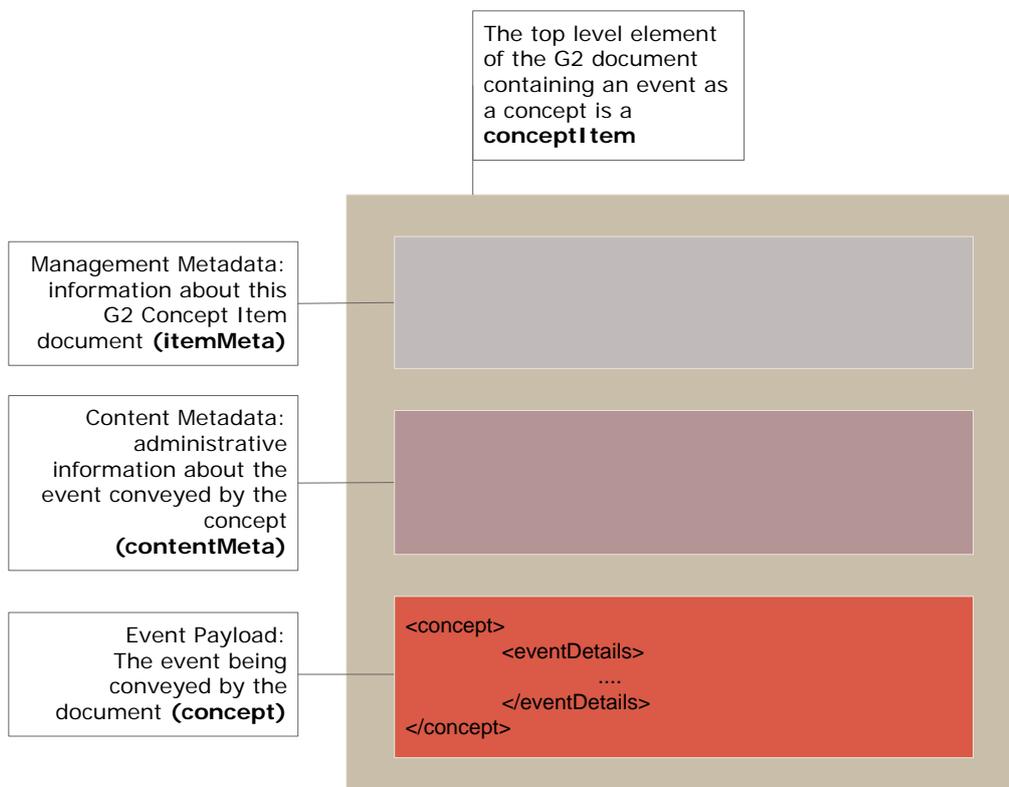


Figure 21: Persistent Event: An event conveyed as a Concept using <eventDetails>

The use-case for the “standalone” implementation would be where 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.

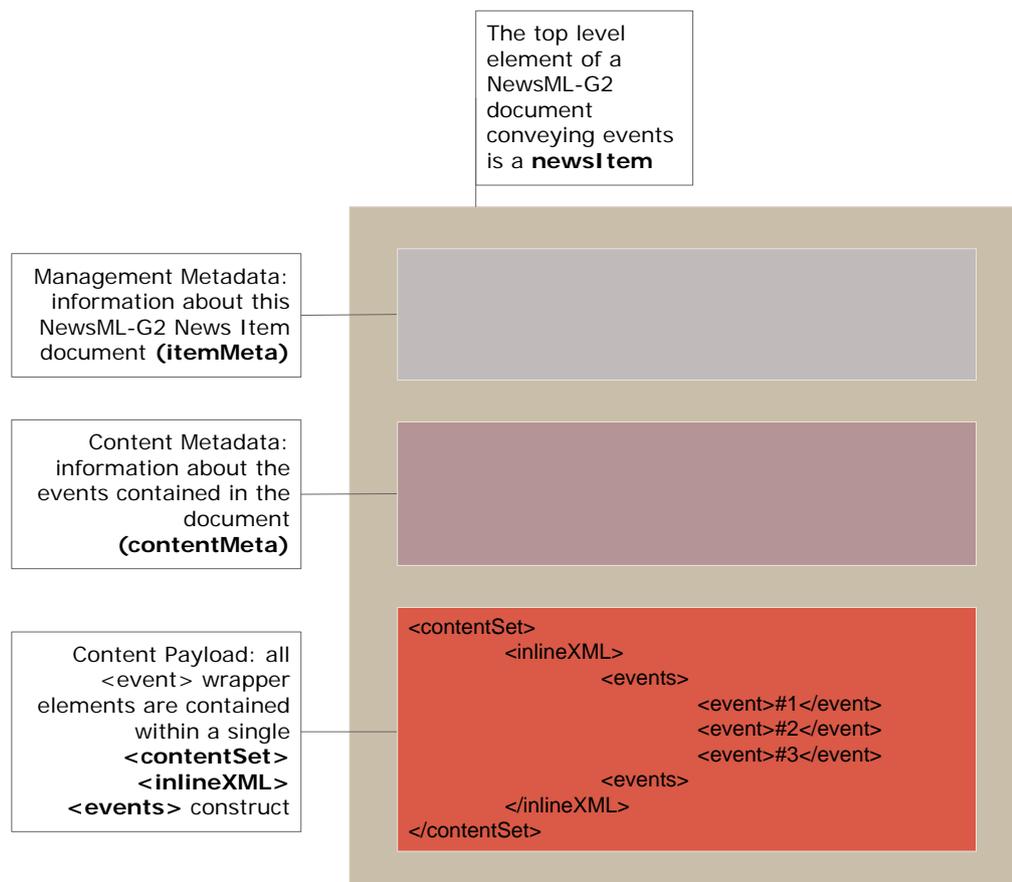


Figure 22: Transient Event(s): Multiple event information carried as <events> in a News Item

14.2 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.

14.2.1 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.

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.

14.2.2 Where does it take place?

The “where” of an event is expressed using the Location property. At the Core Conformance Level (CCL), the <location> element may use a QCode to precisely identify the location from a taxonomy of locations, or a literal value, and may also use a <name> child property to give human-readable details.

The Power level (PCL) gives scope for a rich structure containing detailed information of the event venue (or venues); including GPS coordinates, seating capacity, travel routes etc.

Note that if a PCL structure is used within a document to give in-depth location details, 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.



14.2.3 When does it 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.

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.

14.2.4 Who is involved?

Details of “who” will be present at an event are given using the Participant property. This can be expressed simply and economically at CCL using a QCode or literal value, supplemented by a human-readable Name property. As with the Location property, PCL provides greater power in describing the participants, their roles at the event, and a wealth of related information.

The event organisers are also an important part of the “who” of an event. A set of **Organiser** and **Contact Information** properties can give precise details of the people and organisations responsible for an event, and how to contact them.

14.3 Event Coverage

A vital additional component of an event in a news environment is the need to give information about the intended or actual journalistic response to news events, in terms of personnel or resource assignment, and any content that will be created.

News events may be broadly categorized as:

- ❖ Breaking news – events that are unexpected, considered important, and are likely to quickly develop in complexity and urgency, which need detailed planning and direction of resources.
- ❖ Unplanned news – newsworthy but not necessarily urgent events that are brought to the attention of news organisations during the course of daily operations, for example by a phone call from the public, which need to be added to the news agenda
- ❖ Planned news – events that are known about in advance, such as the 2012 London Olympics, a meeting of the EU Council of Ministers, for which resources need to be scheduled and customers informed.

In all of these cases, 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**).¹⁹

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.

14.4 Event Properties

Employing fictional use-cases, the examples below show the available properties of an event at CCL, 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.

These detailed explanations are followed, in **Event Use Cases**, by a set of detailed examples of events information carried in a News Item and in a Concept Item, illustrating the main differences between the two models.

¹⁹ 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.



14.4.1 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 inflation, amongst other factors. </definition></pre>
<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 literal value <pre><related literal="meeting" /></pre> ❖ a set of attributes to represent some arbitrary value. <pre><related rel="csrel:admission" value="7.0" val ueunit="iso4217:EUR" val uedatatype="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>

14.4.2 Event Relationships

Many news events need to be split into many sub-events in order to manage them effectively. Attempting to convey information and manage coverage for a large event such as the 2012 Summer Olympics as one all-encompassing event would be impractical, if not impossible.

A more logical approach is to break down a very large event into a series of smaller, manageable events, arranged in a hierarchy which can express parent-child relationships as well as simple peer-to-peer relationships.

Using G2, a “master” event is notionally split into sub-events, which in turn may be split into further sub-events without limit. Each event instance may be managed separately yet handled and conveyed within the context of the larger realm of events of which forms a part.

Figure 23 below shows how a hierarchy of events may be created in this way and illustrates the meaning of the <broader>, <narrower>, <related> and <sameAs> relationship properties.

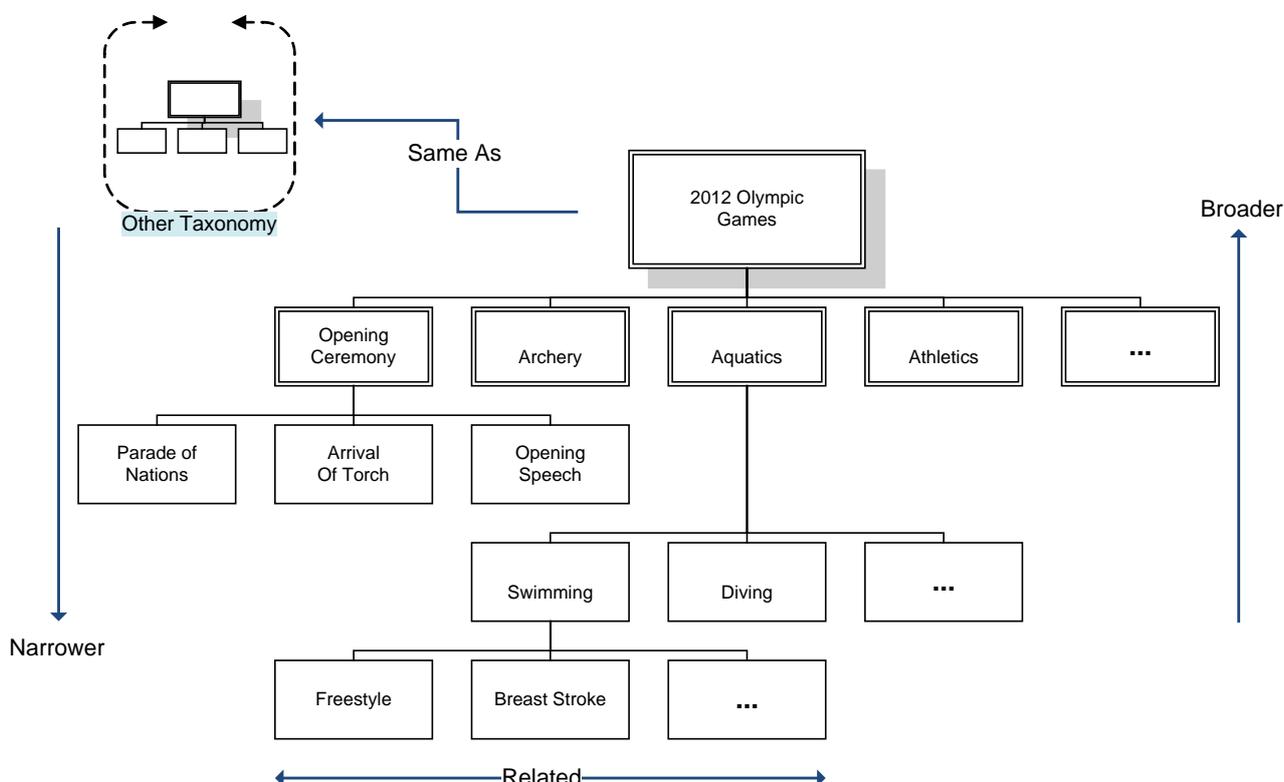


Figure 23: Hierarchy of Events created using Event Relationships

It is important to note from the above diagram that whereas Broader, Narrower and Same As have very specific relationships to the same type of Concept, that is in this case Events, Related has no such restriction. In the diagram:

- ❖ 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 the Same As Swimming or Diving in some other taxonomy.
- ❖ Breast Stroke may be Related to Freestyle, but may also be Related to Diving, Athletics, Parade of Nations, the International Olympic Committee, Michael Phelps, or any other Concept.

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	Repeatable. The event may be part of another event, in which case this



Property	Type	Notes/Example
	(CCL)/Related Concept (PCL)	<p>can be denoted by <broader>.</p> <p>We may use a literal value to identify the broader event that encompasses the current event:</p> <pre><broader type="cpnat: event" literal="int001"> <name>Treasury forecast of economic indicators</name> </broader></pre> <p>But in order to make the related event accessible from the current event, a QCode is needed to reference the related event:</p> <pre><broader type="cpnat: event" qcode="events: 23456789"> <name>Treasury forecast of economic indicators</name> </broader></pre>
<narrower>	Flexible Property/ Related Concept (PCL)	<p>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 event announcing of the committee's decision.</p> <pre><narrower qcode="events: TR2011- 34593"> <name>Minimum Lending Rate announcement press conference</name> </narrower></pre>
<related>	Flexible Property/ Related Concept (PCL)	<p>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 NewsCode:</p> <pre><related rel="irel: seeAlso" type="cpnat: org" qcode="org: fsauk"> <name>UK Financial Services Authority</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 a Government-maintained taxonomy of events:</p> <pre><sameAs qcode="coi: BOE30987"> <name>Monetary Policy Committee of the Bank of England</name> </sameAs></pre>

14.4.3 Event Details Group

The first set of properties of Event Details is date-time information as described in [14.4.3.1](#) below. The further properties are described in [14.4.3.2](#) below

14.4.3.1 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</p>



Property	Type	Notes/Example
		<p>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 data-bbox="579 725 954 860"><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 Schema Duration	<p>Non-repeatable. The time period during which the event takes place is expressed in the form:</p> <p>PnYnMnDnHnMnS</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 data-bbox="579 1921 943 1951"><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>



Property	Type	Notes/Example
		The <dates> wrapper may contain either an <end> date, or a <duration> but not both.
<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>

14.4.3.1.1 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



Property	Type	Notes/Example
		<p>❖ PER SECOND</p> <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 data-bbox="571 488 785 600"><rRule freq="YEARLY" interval="4" /></pre> <p>@until sets a Date with optional Time <i>after</i> which the recurrence rule expires:</p> <pre data-bbox="571 719 858 824"><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 data-bbox="571 943 769 1048"><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> <pre data-bbox="571 1339 1235 1585"><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".</p> <p>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>



Property	Type	Notes/Example
		<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 data-bbox="571 600 890 712"><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 data-bbox="571 786 890 920"><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 data-bbox="571 1122 786 1227"><rRule freq="WEEKLY" wkst="MO" /></pre>
<exDate>	Date with optional Time	<p>Excluded Date of Recurrence. An explicit Date or Dates, with optional Time, excluded from the Recurrence rule. For example, if a regular monthly meeting coincides with public holidays, these can be excluded from the recurrence set using <exDate></p> <pre data-bbox="571 1429 858 1615"><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 data-bbox="571 1816 858 2056"><rRule freq="WEEKLY" until="2011-07-23" /> <rRule freq="WEEKLY" until="2011-12-24" /> <exRule</pre>



Property	Type	Notes/Example
		<pre>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>

14.4.3.2 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> Coverage Intended </name> </newsCoverageStatus></pre>
<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</p> <p>http://cv.iptc.org/newscodes/eventregrole/</p> <p>may be used, which currently has four values:</p> <ul style="list-style-type: none"> ❖ Exhibitor ❖ Media ❖ Public ❖ Student



Property	Type	Notes/Example
		<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="subj: 04000000"> <name xml:lang="en-GB"> Economy, Business and Finance </name> </subject> <subject type="cpnat: abstract" qcode="subj: 04006000"> <name xml:lang="en-GB"> Financial and Business Service </name> </subject> <subject type="cpnat: abstract" qcode="subj: 04006002"> <name xml:lang="en-GB"> Banking </name> </subject></pre>
<location>	Flexible Property/ Flexible Location Property (PCL)	<p>Optional, repeatable property indicating the location of the event.</p> <p>At CCL, a simple QCode or Literal value accompanied by an optional <name> may be used.</p> <pre><location type="cpnat: geoArea" literal="bankofengland"> <name> The Bank of England, Threadneedle Street, London, EC2R 8AH, UK </name> </location></pre>



Property	Type	Notes/Example
		At PCL, a rich Concept-style structure may be used. (See the G2 Specification document for details)
<participant>	Flexible Property/ Flexible Party Property (PCL)	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" <pre><particip ant type="cpnat: person" qcode="pers: 32965"> <name xml: lang="en- GB"> Mervyn King </name> </particip ant></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	Optional, repeatable element for expressing any required conditions for participation in, or attendance at, the event. Either a literal or QCode value may be used. <pre><participati onRequi rement literal="accredi ted"> <name> Accredi tati on requi red </name> </participati onRequi rement></pre>
<organiser>	Flexible Property/ Flexible Party Property (PCL)	Optional, repeatable. Describes the organiser of the event. <pre><organi ser type="cpnat: org" literal="iptc"> <name xml: lang="en- GB"> Internati onal Press Tel ecommuni cations Council </name> <name xml- lang="fr"> Comi té Internati onal de Tél écommuni cations de Presse </name> </organi ser></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 14.4.3.2.1 below



Property	Type	Notes/Example
<language>	-	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. <pre><language tag="en"> <name>English</name> </language></pre>
<newsCoverage>		DEPRECATED, Should not be used with Items conforming to EventsML-G2 1.6 and later. News Coverage is now part of the G2 Planning Item. (See Editorial Planning – the Planning Item)

14.4.3.2.1 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	An "Electronic Address" type allows the expression of @role (QCode) to qualify the information, for example <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: <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 14.4.3.2.1.1 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.



Property	Type	Notes/Example
<note>	Block	Any other contact-related information, such as “Office closed for lunch from 12.30pm for three hours”

14.4.3.2.1.1 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 either a QCode or Literal value with optional <name> child element
<area>	Flexible Property	For example, name of a county or region. May be either a QCode or Literal value with optional <name> child element
<country>	Flexible Property	The country name may be either a QCode referring to a CV of countries, or Literal value with optional <name> child element
<postalCode>	Internationalized String	The postal code, zip code or equivalent

For example:

```
<address role="address:registered">
  <line>20 Garrick Street</line>
  <locality literal="London" />
  <country qcode="ISOCountryCode:UK">
    <name xml:lang="en">United Kingdom</name>
  </country>
  <postalCode>WC2E 9BT</postalCode>
</address>
```

14.5 Event Use Cases

Some news events may be regarded as transient: when the event has taken place, the event information (as opposed to any content generated) will not be used or referred to again. In contrast, other events need to be persisted in a knowledge base so that they can be updated and referenced over time.

The G2 family supports both use cases, as detailed [Structure and compatibility](#). Events that are referenced over time are conveyed using Concept Items and Knowledge Items; this method ensures each event carries a unique Concept ID. Transient event information may be conveyed by News Items.

Figure 24 shows how the same information about an event may be conveyed by the different wrapper elements within a News Item (for a transient event) and a Concept Item (for persistent events).

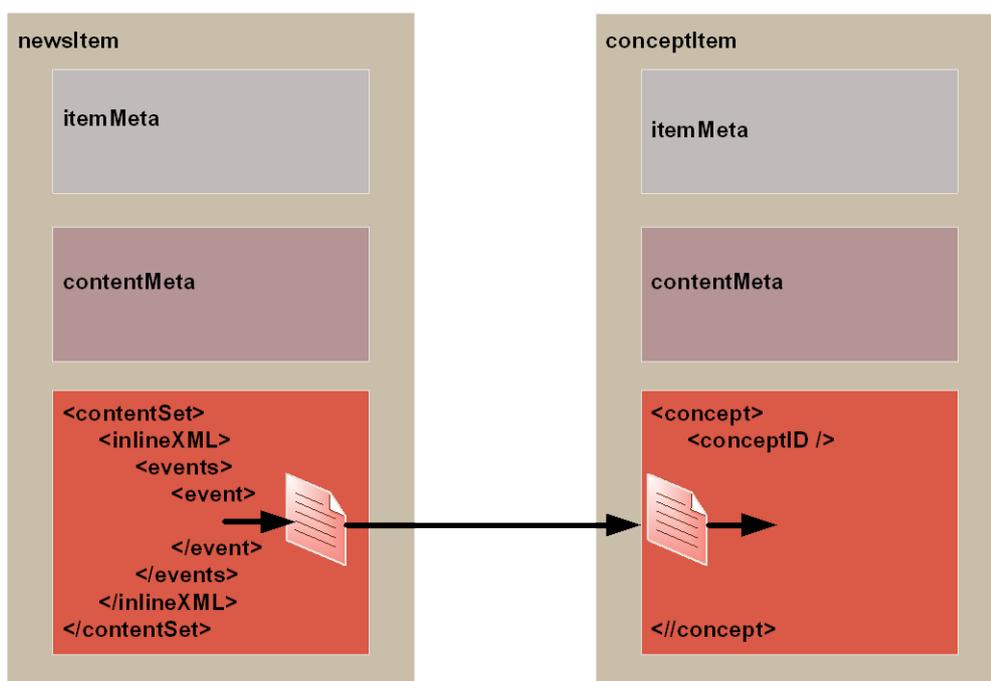


Figure 24: An event represented in a News Item or a Concept Item uses a common structure

The major difference between the two implementations is that the event carried within a Concept Item is persistent; it has a Concept ID (a QCode) which must be globally unique and thus can be used by other Items to reference the event. The information about individual events carried within a News Item is transient, and has no means of being referenced outside the scope of the News Item that carries it.

The business case for transient events in a News Item would be that the provider needs to send a collection of events, whether related or not, that will be neither referenced in content nor separately managed by the provider. The events could be managed independently by the receiver, or they could be intended for use as a “snapshot” of upcoming events, such “a diary of events for the coming week/month/year”.

Event Concepts may also be collected and conveyed as a single Item:

- ❖ as one or more *Concept components* in the Concept Set of a *Knowledge Item*: Administrative Metadata for each of the Concept components in a KI is available using `<partMeta>`. These collections can be physically sent, or hosted and made available online. The Event Concepts in a Knowledge Item do not have to be related; this method enables all events to be conveyed in one container.
- ❖ as a collection of *references to Concept Items* in the Group Set of a *Package Item*: This enables providers to express more comprehensive management information about each Event Concept Item.

14.5.1 Event in a Concept Item

The following code snippets show how a news event expressed as a Concept is 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.



The `<conceptItem>` `@guid` and `@version` uniquely identify a given CI, for the purpose of managing and updating the event information contained in the `<concept>` wrapper. Items that reference the event MUST do so ONLY by using the Concept ID. Why? Because the Concept ID uniquely references a persistent Web resource, whereas the GUID only identifies a document containing the information, and applications/consumers may have only retained the `<concept>` (and its Concept ID), but not the `<conceptItem>` (and its GUID).



To enable concepts to be identified by a Concept ID QCode, a reference to the provider's catalog (or a catalog statement giving the scheme URI) MUST be included:

```
<?xml version="1.0" encoding="UTF-8"?>
<conceptItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  guid="urn:newsml:iptc.org:20110126:qqwpi ruuew4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
</conceptItem>
```

Annotations:

- Top Level element (points to <conceptItem>)
- XML Schema document GUID and version NAR standard, version and conformance (points to XSD/NewsML-G2_2.9-spec-All-Power.xsd)
- Provider's catalog (points to http://www.example.com/events/event-catalog.xml)

The <itemMeta> and <contentMeta> blocks (if applicable) are common to all G2 items. For a Concept Item, 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="cinat:concept" />
  <providerLiteral="IPTC" />
  <versionCreated>2011-01-26T14:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
<urgency>5</urgency>
<contentCreated>2011-01-26T12:15:00Z</contentCreated>
<contentModified>2011-01-26T13:35:00Z</contentModified>
</contentMeta>
```

Annotations:

- Concept Item nature (points to <itemClass qcode="cinat:concept" />)

The content of a Concept Item 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 being conveyed. The type of concept being conveyed is expressed using the <type> property 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/>.

```
<concept>
  <conceptId created="2011-01-26T12:00:00Z" qcode="event:1234567" />
  <type qcode="cpnat:event" />
</concept>
```

Annotations:

- This is an Event (points to <type qcode="cpnat:event" />)

There is NO <events> and <event> wrapper elements, the event information, such as <name> and <eventDetails> is placed directly within the <concept> wrapper.

```
<concept>
  <conceptId created="2011-01-26T12:00:00Z" qcode="event:1234567" />
  <type qcode="cpnat:event" />
  <name>IPTC Autumn Meeting</name>
  <eventDetails>
    <dates>
      <start>2011-10-15T09:00:00+02:00</start>
      <duration>P3D</duration>
    </dates>
    . . . . .
  </eventDetails>
</concept>
```

In the code listings below, we show how an event can be sent in the first instance, giving outline details, and subsequently updated by re-sending information using the same GUID.

LISTING 16 Event sent as a Concept Item

```
<?xml version="1.0" encoding="UTF-8"?>
<conceptItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
```



```

XSD/NewsML-G2_2.9-spec-All-Power.xsd"
guid="urn:newsml:iptc.org:20090122:qqwpi ruuew4711"
version="2"
standard="NewsML-G2"
standardversion="2.9"
conformance="power"
xml:lang="en">
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
<catalogRef href="http://www.example.com/events/event-catalog.xml" />
<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider literal="IPTC" />
  <versionCreated>2010-10-18T13:05:00Z
</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
<urgency>5</urgency>
<contentCreated>2010-10-16T12:15:00Z</contentCreated>
<contentModified>2010-10-16T13:35:00Z</contentModified>
</contentMeta>
<concept>
  <conceptId created="2010-10-16T12:15:00Z" qcode="event:1234567" />
  <type qcode="cpnat:event" />
  <name>IPTC Autumn Meeting</name>
  <eventDetails>
    <dates>
      <start>2010-11-03T09:00:00+01:00</start>
      <duration>P3D</duration>
    </dates>
    <location literal="Hotel Raffaello">
      <name>Hotel Raffaello, Via Urbana, 3 - 00184 Rome, Italy</name>
      <related rel="frel:venue" qcode="ventyp:hotel" />
      <POIDetails>
        <position latitude="41.897917" longitude="12.495511" />
        <contactInfo>
          <web>http://www.hotelraffaello.it</web>
        </contactInfo>
      </POIDetails>
    </location>
  </eventDetails>
</concept>
</conceptItem>

```

[<return to Code Listings index](#)

After some passage of time, the provider may have learned of changes to details of the event, and will send an update to the customer using a Concept Item with the same GUID, but an incremented version number:

```

<?xml version="1.0" encoding="UTF-8"?>
<conceptItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:iptc.org:20110126:qqwpi ruuew4711" ← same GUID
  version="2" ← new version number
  standard="NewsML-G2"

```

Timestamps used in the document will also be updated to show that this is newer than the previous version.

We could also provide a natural language <edNote> to inform receivers why the update has been sent, and a machine-readable <signal> to give a processing instruction.

```

<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider literal="IPTC" />
  <versionCreated>2011-01-26T14:45:00Z ← Item timestamp
</versionCreated>
  <pubStatus qcode="stat:usable" />
  <edNote>Venue changed by the organisers</edNote> ← Editorial note
  <signal qcode="eventproc:replace"> ← Signal to processor
    <name>Replace existing</name> ← Natural language instruction
  </signal>
</itemMeta>
<contentMeta>
<urgency>5</urgency>

```



```
<contentCreated>2011-01-26T12:15:00Z</contentCreated>
<contentModified>2011-01-26T14:35:00Z</contentModified> ← Concept timestamp
```

In the following listing, the complete event information is re-sent, with a processing instruction that the previous version should be replaced.

LISTING 17 Update to the Previously-Sent Event

```
<?xml version="1.0" encoding="UTF-8"?>
<conceptItem
  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/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  guid="urn:newsm1:iptc.org:20090122:qqwpi ruuew4711"
  version="3"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T14:05:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
    <edNote>Updated - conference venue has changed! </edNote>
    <signal qcode="eventproc:replace">
      <name>Replace existing</name>
    </signal>
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2010-10-18T13:15:00Z</contentCreated>
    <contentModified>2010-10-18T13:35:00Z</contentModified>
  </contentMeta>
  <concept>
    <conceptId created="2010-10-16T12:15:00Z" qcode="event:1234567" />
    <type qcode="cpnat:event" />
    <name>IPTC Autumn Meeting</name>
    <eventDetails>
      <dates>
        <start>2010-11-03T09:00:00+01:00</start>
        <duration>P3D</duration>
      </dates>
      <location literal="Residenza di Ripetta">
        <name>Residenza di Ripetta, via di Ripetta,
          231 - 00186 Roma - Italy</name>
        <related rel="frel:venue" qcode="ventyp:hotel" />
        <POIDetails>
          <position latitude="41.908617" longitude="12.47592" />
          <contactInfo>
            <web>http://www.ripetta.it</web>
          </contactInfo>
        </POIDetails>
      </location>
    </eventDetails>
  </concept>
</conceptItem>
```

[<<return to Code Listings index](#)

14.5.2 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 element is <knowledgeItem>, which contains identification, version and catalog information, in common with other Items.

```
<?xml version="1.0" encoding="UTF-8"?>
<knowledgeItem ← Top level 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/
XSD/NewsML-G2_2.9-spec-All-Power.xsd"
guid="urn:newsml:iptc.org:20110126:qqwpi ruuew4712"
version="1"
standard="NewsML-G2"
standardversion="2.9"
conformance="power"
xml:lang="en"
<catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
<catalogRef href="http://www.example.com/events/event-catalog.xml" />
```

Document GUID
 and version
 standard,
 standard version
 and conformance

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 literal="IPTC" />
  <versionCreated>2011-01-26T12:00:00Z</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
  <urgency>5</urgency>
  <contentCreated>2011-01-26T12:15:00Z</contentCreated>
  <contentModified>2011-01-26T14:35:00Z</contentModified>
  <subject qcode="subj:04010000">
    <name>media</name>
  </subject>
  <subject qcode="subj:04010004">
    <name>news agency</name>
  </subject>
  <subject qcode="subj:13022000">
    <name>IT/computer sciences</name>
  </subject>
</contentMeta>
```

payload a concept
 document timestamp
 Item status
 Content
 timestamps
 Subject codes
 of all the events

The two events in the listing are related, with the relationship indicated by the second event using the <broader> property to show that it is an event which is part of the three-day event listed first

First event:

```
<conceptSet>
  <concept>
    <!-- FIRST EVENT! -->
    <!-- x -->
    <conceptId
      created="2010-10-30T12:00:00+00:00"
      qcode="event:1234567"
    />
    <name>IPTC Autumn Meeting</name>
    <eventDetails>
      ....
```

Concept ID

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! -->
  <!-- x -->
  <conceptId
    created="2010-10-30T12:00:00+00:00"
    qcode="event:91011123"
  />
  <name> Presentation of ODF (Olympic Data Format) by Elena Garrido</name>
  <broader
    type="cinat:event"
    qcode="event:1234567">
    <name>IPTC Autumn Meeting</name>
  </broader>
  <eventDetails>
```

Ref to Concept ID
 of first event



Although Knowledge Items are a convenient way to convey a set of related events, there is no requirement that all of the events 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 18 Two Related Events conveyed in a Knowledge Item

```
<?xml version="1.0" encoding="UTF-8"?>
<knowledgitem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  guid="urn:newsml:iptc.org:20101019:qqwpi ruuew4712"
  version="1"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-26T12:00:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2009-01-26T12:15:00Z</contentCreated>
    <contentModified>2009-01-26T14:35:00Z</contentModified>
    <subject qcode="subj:04010000">
      <name>media</name>
    </subject>
    <subject qcode="subj:04010004">
      <name>news agency</name>
    </subject>
    <subject qcode="subj:13022000">
      <name>IT/computer sciences</name>
    </subject>
  </contentMeta>
  <conceptSet>
    <concept>
      <!-- FIRST EVENT! -->
      <!-- X -->
      <conceptId created="2010-10-16T12:15:00Z" qcode="event:1234567" />
      <name>IPTC Autumn Meeting</name>
      <eventDetails>
        <dates>
          <start>2010-11-03T09:00:00+01:00</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 31 October 2010
        </registration>
        <participationRequirement>
          <name>Membershi p</name>
          <definition>Only members of the IPTC and their invited
            guests may attend.
          </definition>
        </participationRequirement>
        <accessStatus qcode="accst:easy" />
        <language tag="en" />
        <organiser literal="IPTC" role="orgrole:mainorganiser">
          <name>International Press Telecommunications Council</name>
          <organisationDetails>
            <founded>1965</founded>
          </organisationDetails>
        </organiser>
        <contactInfo>
          <email>mdirector@i p c t . o r g</email>
          <note>Mi chael Stei dl, Managi ng Di rector</note>
          <web>http://www.i p t c . o r g</web>
        </contactInfo>
        <location literal="Residenzadi Ripetta">
          <name>
            Residenza di Ripetta, via di Ripetta, 231 - 00186 Roma - Italy
          </name>
        </location>
      </eventDetails>
    </concept>
  </conceptSet>
</knowledgitem>
```



```

</name>
  <related rel = "frel:venue" qcode = "ventyp:hotel" />
  <POIDetails>
    <position latitude = "41.908617" longitude = "12.47592" />
    <contactInfo>
      <web>http://www.ripetta.it</web>
    </contactInfo>
  </POIDetails>
</location>
<participant literal = "StéphaneGuérillot">
  <name>Stéphane Guérillot</name>
  <definition role = "drol:jobtitle">IPTC Chairman</definition>
</participant>
<participant literal = "MichaelSteidl">
  <name>Michael Steidl</name>
  <definition role = "drol:jobtitle">Managing Director</definition>
</participant>
</eventDetails>
</concept>
<concept>
  <!-- SECOND EVENT! -->
  <!-- x -->
  <conceptId created = "2000-10-30T12:00:00+00:00" qcode = "event:9101123" />
  <name>Presentation of ODF (Olympic Data Format) by Elena Garrido</name>
  <broader type = "cpnat:event" qcode = "event:1234567">
    <name>IPTC Autumn Meeting</name>
  </broader>
  <eventDetails>
    <dates>
      <start>2009-11-03T14:00:00+01:00</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 literal = "ElenaGarrido">
      <name>Elena Garrido</name>
      <definition role = "drol:sessionrole">Presenter</definition>
    </participant>
    <participant literal = "MichaelSteidl">
      <name>Michael Steidl</name>
      <definition role = "drol:sessionrole">Moderator</definition>
    </participant>
  </eventDetails>
</concept>
</conceptSet>
</knowledgeItem>

```

[<<return to Code Listings index](#)

When multiple event Concepts are conveyed or hosted in a Knowledge Item, and a sub-set of the Concepts is updated, providers may need to inform customers WHAT has changed in the new version of the Knowledge Item.

This information can be indicated using a <partMeta> wrapper which contains references to parts of the Item's content. The referencing mechanism uses the standard XML ID/IDREF mechanism. The snippet below shows that the Concept with @id="conceptA" (the ID) was modified on '2010-09-15', as implied by the Part Meta with @contentRefs="conceptA" (the IDREF):

```

<?xml version="1.0" encoding="UTF-8"?>
<knowledgeItem
  xmlns="..."
  <catalogRef ... />
  <itemMeta>
    ...
  </itemMeta>
  <contentMeta>
    ...
  </contentMeta>
  <partMeta contentRefs="conceptA">
    <contentModified>2010-09-15</contentModified>
  </partMeta>

```

```

</partMeta>
<conceptSet>
  <concept id="conceptA">
    <!-- FIRST EVENT! -->
    <!-- x -->
    <conceptId created="2010-08-30T12:00:00+00:00" qcode="event:1234567" />
    ...
  </concept>
  <concept>
    <!-- SECOND EVENT! -->
    <!-- x -->
    <conceptId created="2010-08-30T12:00:00+00:00" qcode="event:91011123" />
    ...
  </concept>
</conceptSet>
</knowledgeItem>

```

14.5.3 Packaging Events using <conceptRef>

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. As managed and persistent events, they would be identified using QCodes. These News Events can be collected and conveyed within a Package <group> using a Concept Reference <conceptRef> as follows:

```

<groupSet root="G1">
  <group id="G1" role="group:main">
    ...
    <conceptRef type="cinat:event" qcode="iptcevents:20081007135637.12">
      <name>Barack Obama arrive à Washington</name>
    </conceptRef>
    ...
  </group>
</groupSet>

```

(See **Packages** chapter for details of the Package Item structure and properties)

The datatype of a Concept Reference at CCL is a Flexible Property (FlexPropType) as illustrated above. The optional @type indicates that the concept is an Event, and the @qcode is a unique identifier that can be de-referenced to get further information about the event. The optional <name> child element is a natural-language name for the event.

At PCL, further properties as defined by the Flex1ConceptPropType may also be given, however, the provider must take care that any additional properties do not refine or re-define the original Concept, as this persists in the original Item. Changes to the original Concept MUST be conveyed in a Concept Item or Knowledge Item.



The standard method of exchanging concept information is the NewsML-G2 Knowledge Item, which is a container for concepts and the detailed properties of each concept being conveyed. Why use a Package to convey concepts, and not a Knowledge Item? The roles of both Items are illustrated by the use case below.

14.5.3.1 Use Case for a Package of News Events

A news organisation sends its customers a news events service. The event information is expressed using EventsML-G2 properties and each Event is identified using a Concept ID (QCode). Customers use the events service as part of their own editorial diary, or daybook. The event information is published using NewsML-G2 Knowledge Items, where each KI is a set of concepts identified by QCodes and containing the facts about each event as EventsML-G2 properties.

The news organisation also provides a Top Events of the Day service. Its editors choose the most significant news events of the day and package them in an ordered list to be sent to the customers.

The Events Package has the following characteristics:

- ❖ It is an editorial product. As such, the value of the package is in the journalistic judgement used to select the events, not in the compilation of the event information itself.
- ❖ It is lightweight, containing only a reference to 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.

- ❖ It can be ordered to indicate the relative significance of the events, but no other relationship between the events is expressed or implied.
- ❖ New versions of the Package may be published; events can be replaced or deleted from the package, but details of the events themselves cannot be changed.

In contrast, 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 should not be replaced or deleted, as this breaks existing references to the concepts.

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.

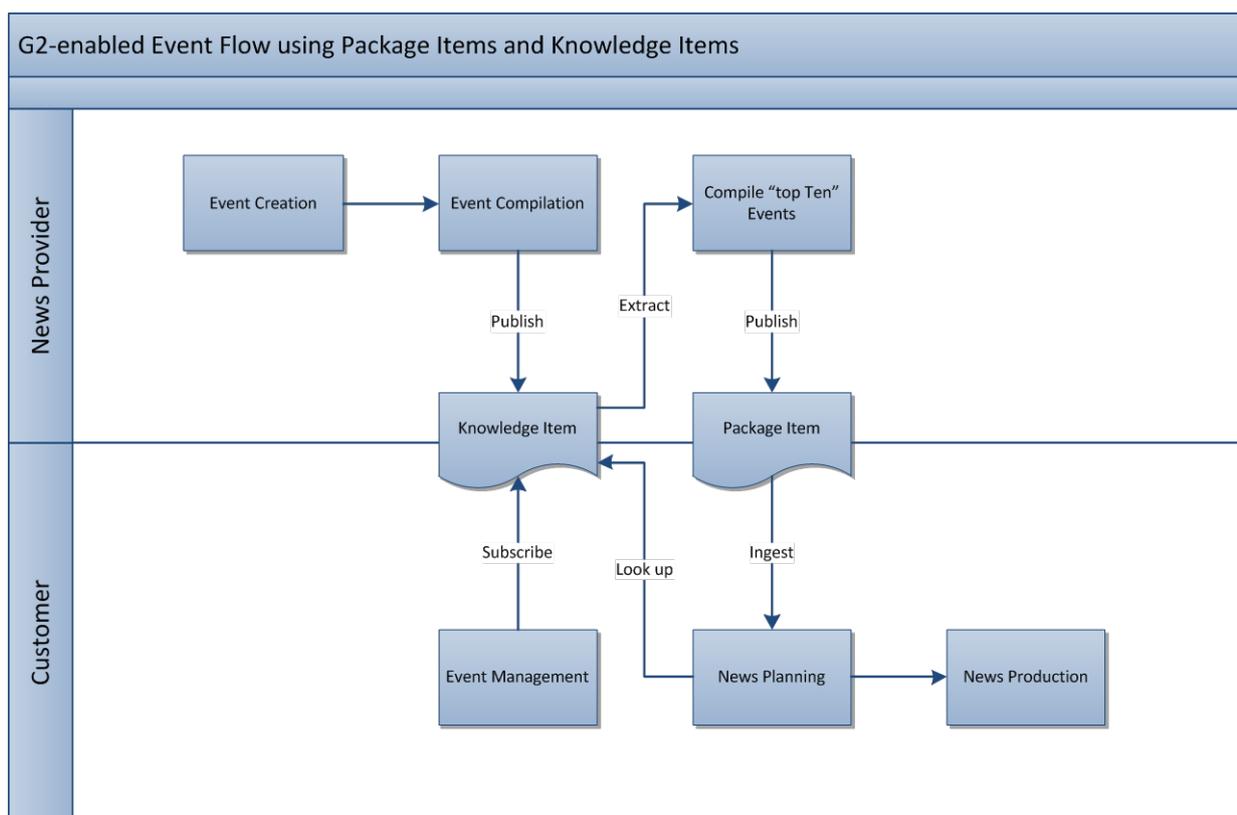


Figure 25: Event Flow using Package Items and Knowledge Items

14.5.4 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.

One of the available child elements of the Content Set of a News Item is the <inlineXML> wrapper element, which can convey any valid XML content – in this case an <events> wrapper containing one or more <event> instances, each of which is a separate self-contained set of event information.

Using this method, we can convey the event listing inside a NewsML-G2 <newsItem> containing id, versioning, schema and catalog information:

```
<?xml version="1.0" encoding="UTF-8"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  guid="urn:newsml:iptc.org:20110122:qqwpi ruuew4711"
  version="1" standard="NewsML-G2" standardversion="2.9" conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
```

The <itemMeta> block holds management information about the NewsML-G2 document (NOT the events conveyed by the document, as these event definitions are transient). The Item Class has a value from the IPTC News Item Nature (<http://cv.iptc.org/newscodes/nature/>) NewsCode of "composite", together with Provider, a Time Stamp and Publishing Status

```
<itemMeta>
  <itemClass qcode="ni nat: composite" />
  <provider literal="IPTC" />
  <versionCreated>2011-01-22T12:00:00Z</versionCreated>
  <pubStatus qcode="stat: usable" />
</itemMeta>
```

The **Administrative Metadata** group of properties of the <contentMeta> block may be used, but in this case, the events are not related and do not share any property values of significance to a receiver, so <contentMeta> may be omitted.

The <contentSet> block may have child elements chosen from either <inlineXML>, <inlineData> or <remoteContent>. In this case we will be using <inlineXML> to wrap an <events> structure. (Event information has been abridged in the listing for ease of reading.)

LISTING 19 A Set of Events carried in a News Item

```
<?xml version="1.0" encoding="UTF-8"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  guid="urn:newsml:iptc.org:20090122:qqwpi ruuew4711"
  version="2" standard="NewsML-G2" standardversion="2.9" conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <itemMeta>
    <itemClass qcode="ni nat: composite" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T12:44:00Z</versionCreated>
    <pubStatus qcode="stat: usable" />
  </itemMeta>
  <contentSet>
    <inlineXML>
      <events>
        <!-- x -->
        <!-- FIRST EVENT! -->
        <!-- x -->
        <event>
          <name>IPTC Autumn Meeting</name>
          <eventDetails>
            <dates>
              <start>2010-11-03T09:00:00+02:00</start>
              <duration>P3D</duration>
            </dates>
            <location literal="Residenzadi Ripetta">
              <name>
                Residenza di Ripetta, via di Ripetta, 231 - 00186 Roma - Italy
```



```

        </name>
        <related rel="frel:venueType" qcode="ventyp:hotel" />
        <POIDetails>
            <position latitude="41.908617" longitude="12.47592" />
            <contactInfo>
                <web>http://www.ripetta.it</web>
            </contactInfo>
        </POIDetails>
    </location>
</eventDetails>
</event>
<!--          x          -->
<!--          SECOND EVENT!          -->
<!--          x          -->
    <event>
        <name>Presentation of the Olympic Data Format (ODF)</name>
        <eventDetails>
            <dates>
                <start>2010-11-03T14:00:00+01:00</start>
                <end>2010-11-03T15:00:00+01:00</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="facilncd:Food" />
        <related rel="frel:facility" qcode="facilncd:Ai rCondi ti oning" />
        <eventDetails>
            <dates>
                <start>2011-02-03T19:30:00+01:00</start>
                <end>2011-02-05</end>
                <rRule freq="DAILY" byday="TH FR SA" />
            </dates>
        </eventDetails>
    </event>
</events>
</inlineXML>
</contentSet>
</newsItem>

```

[<<return to Code Listings index](#)

14.5.5 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 18** an event was conveyed as a concept within the Concept Set of a Knowledge Item:

```

<knowledgeItem>
    ...
    <conceptSet>
        <concept>
            <!--          FIRST EVENT!          -->
            <!--          -->
            <conceptId created="2010-10-30T12:00:00+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>` property, as shown below:

```
<!-- available at Core Conformance Level -->
<newsItem conformance="core" ... >
  ...
  <contentMeta>
    ...
    <subject type="cpnat: event" qcode="event: 12345657">
      <name>IPTC Autumn Meeting</name>
    </subject>
    ...
  </contentMeta>
  ...
</newsItem>
```

Referencing an event using the subject property 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.

14.5.5.1 Using `<bag>` to create a composite concept

At the Power Conformance Level, a more advanced construct enables implementers to express a richer relationship to the event and entities related to the event, using a `<bag>` child element of `<subject>`. The `<bag>` is used to hold a collection of concepts that together create a *new*, composite, concept.

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="fi nevent: takeover123AB" />
```

... and the companies involved are also members of a CV of companies. Thus the provider could also add subject properties for each company:

```
<subject type="cpnat: organi sati on" qcode="i si n: Smal l Company" />
<subject type="cpnat: organi sati on" qcode="i si n: Gl obal Company" />
```

However, a new, richer, composite (abstract) concept combining the concept IDs for the Event *and* the two companies could be created instead using `<bag>`:

```
<-- feature available at Power Conformance Level -->
<newsItem conformance="power" ... >
  ...
  <contentMeta>
    ...
    <subject type="cpnat: abstract">
      <name>Gl obal Company takes over Smal l Company</name>
      <bag>
        <bit type="cpnat: event" qcode="fi nevent: takeover123AB" />
        <bit type="cpnat: organi sati on" qcode="i si n: Smal l Company" />
        <bit type="cpnat: organi sati on" qcode="i si n: Gl obal Company" />
      </bag>
    </subject>
    ...
  </contentMeta>
  ...
</newsItem>
```



Note that in this case, the `<subject>` does *not* have its own `@qcode` or `@literal` identifier; these properties must NOT be used when `<subject>` has a `<bag>` element

14.5.5.2 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 use-case is that 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="fi nevent: takeover123AB" />  
    <bit type="cpnat: organi sati on" qcode="i si n: Smal l Company" si gni fi cance="100" />  
    <bit type="cpnat: organi sati on" qcode="i si n: Gl obal Company" si gni fi cance="10" />  
  </bag>  
</subj ect>
```



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>.



This page intentionally blank



15 Editorial Planning – the Planning Item

The news-gathering process is not an ad-hoc process. As outlined in [Chapter 5 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, a significant new addition to the G2 family²⁰, is designed to address this need. As a “top level” NewsML-G2 Item, it joins the family 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 [22.2](#) and [22.3](#) 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>.

15.1 Item Metadata <itemMeta>

The standard properties of <itemMeta> (see [Item Metadata](#)) 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" />
```

²⁰ Available from NewsML-G2 v 2.7 and EventsML-G2 v1.6 and up,

15.2 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, (7.5.1.2) may be used. The Core Group is: <language>, <keyword>, <subject>, <slugline>, <headline>, <description>

15.3 <newsCoverageSet>

The <newsCoverageSet> wraps one or more <newsCoverage> components (see Figure 26). 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.

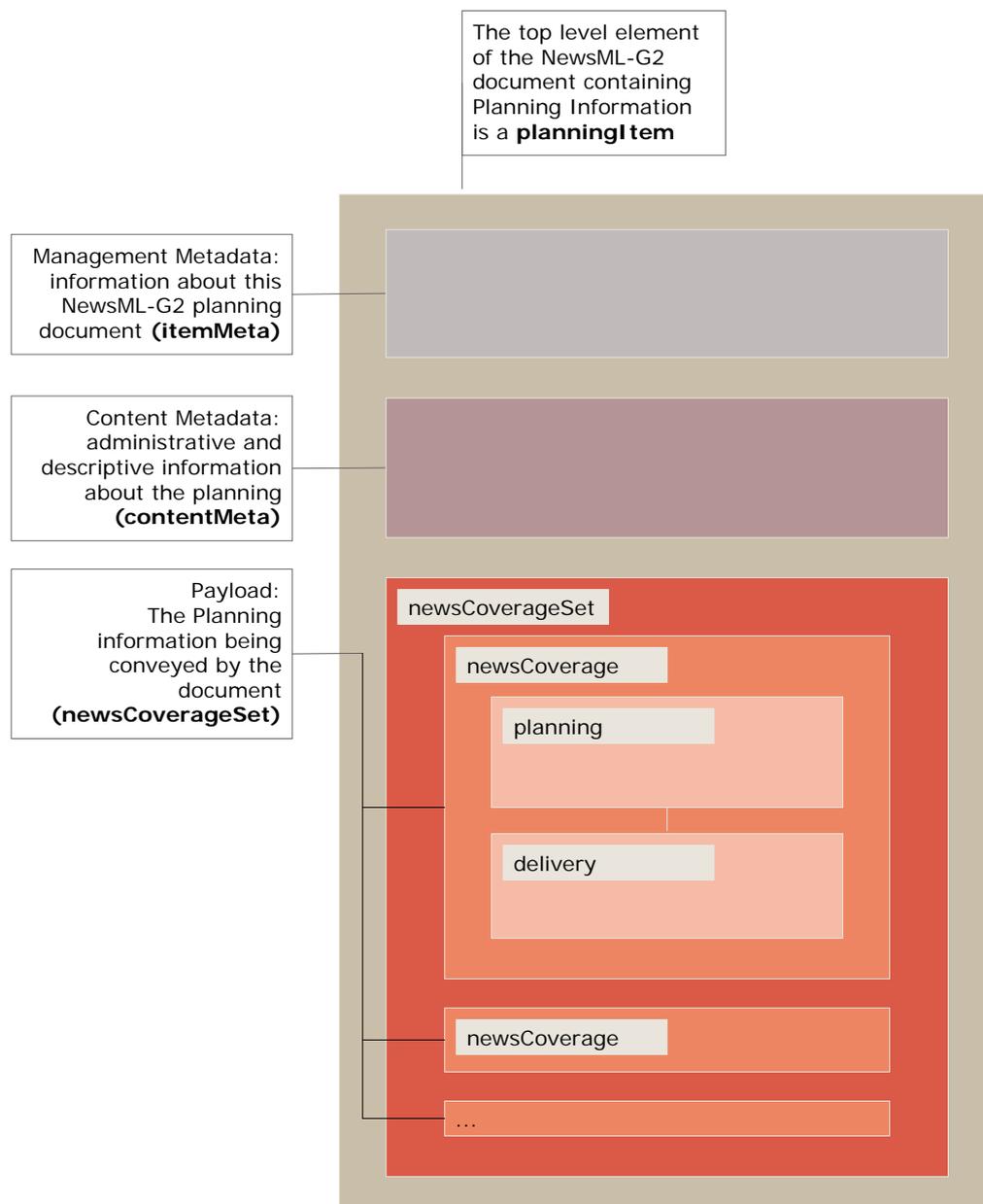


Figure 26: Planning Item: the newsCoverageSet has one or more newsCoverage components

15.4 <newsCoverage>

Each <newsCoverage> component MUST contain at least one <planning> container and optionally a <delivery> container.



15.5 <planning>

At CCL, planning information is expressed as a Block type <edNote> property, with some mark-up permitted. At PCL, more extensive, detailed information may be given as discussed below.

Property	Type	Notes/Example
<edNote>	Block	Optional, repeatable. Editorial note giving natural language information on the planned coverage <edNote>Additional media release expected</edNote>

LISTING 20 Planning Item at CCL

```
<?xml version="1.0" encoding="UTF-8"?>
<planningItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd"
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="core"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="plinat:newscoverage" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T13:05:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2010-10-16T12:15:00Z</contentCreated>
    <contentModified>2010-10-16T13:35:00Z</contentModified>
  </contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2008-01-26T13:19:11Z">
      <planning>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2009-01-26T15:19:11Z">
      <planning>
        <edNote>Picture scheduled 2009-11-06T13:00:00+02:00</edNote>
      </planning>
      <delivery>
        <deliveredItemRef
          guidref="tag:example.com,2008:ART-ENT-SRV:USA20081220098658">
        </deliveredItemRef>
      </delivery>
    </newsCoverage>
  </newsCoverageSet>
</planningItem>
```

[<return to Code Listings index](#)

15.5.1 <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.

Thus, 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.

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 content to be delivered is a NewsML-G2 News Item. <pre><g2ContentType> appl icat ion/vnd. iptc. g2. newsi tem+xml </g2ContentType></pre>
<itemClass>	QCode	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” <pre><i temCl ass qcode=“ni nat: text” /></pre>
<itemCount>	Empty	The number of items to be delivered, expressed as a range: <pre><i temCount 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	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: <pre><schedul ed>2009- 10- 16</schedul ed></pre>
<service>	Qualified Property	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. <pre><servi ce qcode=“srv: i ntwire”> <name>Internati onal Wi re Servi ce</name> </servi ce></pre>
<assignedTo>	Flex1 Party Property	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><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”</pre>



Property	Type	Notes/Example
		<pre>qcode="pers: 54321" > <name>Stilton Cheesewright</name> <personDetails> <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>

15.5.1.1 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	Optional, repeatable. Natural language author/creator information: <by>By Stilton Cheesewright</by>
<creditline>	String	Optional, repeatable. A freeform expression of the credit(s) for the content: <creditline>Additional reporting by Bertram Wooster</creditline>
<dateline>	Label	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: <dateline>Totley Towers, January 27, 2009 (Reuters)</dateline>
<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": <description role="drol:summary">Stilton Cheesewright will report on the proceedings from the NewsCodes Working Party</description>
<genre>	Flex1 Concept Property	Optional, repeatable. The nature of the journalistic content that is intended for the news coverage. May be expressed by a Literal or QCode value, with optional @type. Child elements may be any from the Concept Definition Group and the Concept Relationships Group (see Concepts)



Property	Type	Notes/Example
		<p>and Concept Items):</p> <pre><genre qcode="mygenre: main"> <name>Main article </name> </genre></pre>
<headline>	Label	<p>Optional, repeatable. Headline that will apply to the content:</p> <pre><headline>NewsCodes Working Party</headline></pre>
<keyword>	String	<p>Optional, repeatable. A freeform term to assist in indexing the content:</p> <pre><keyword>IPTC</keyword></pre>
<language>	-	<p>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.</p> <p>The language @tag MUST be expressed using IETF BCP 47 and may have a child element of <name>:</p> <pre><language tag="en-GB"> <name>UK English</name> </language></pre>
<slugline>	Internationalised String	<p>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:</p> <pre><slugline separator="-">US-AUTO-BAILOUT</slugline></pre>
<subject>	Flex1 Concept Property	<p>Optional, repeatable. Indicates the subject matter of the intended coverage.</p> <p>Child elements may be any from the Concept Definition Group and the Concept Relationships Group (see Concepts and Concept Items):</p> <pre><subject qcode="subj:04010000"> <name>media</name> subject> <subject qcode="subj:04010004"> <name>news agency</name> </subject> <subject qcode="subj:13022000"> <name>IT/computer sciences</name> </subject></pre> <p>Or</p> <pre><subject qcode="subj:04010004"> <name>news agency</name> <broader qcode="subj:04010000"> <name>Media</name> </broader> </subject> <subject qcode="subj:13022000"> <name>IT/computer sciences</name> </subject></pre>



LISTING 21 Planning Item at PCL

```
<?xml version="1.0" encoding="UTF-8"?>
<planningItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="plinat:newscoverage" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T13:05:00Z
    </versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2010-10-16T12:15:00Z</contentCreated>
    <contentModified>2010-10-16T13:35:00Z</contentModified>
  </contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2008-01-26T13:19:11Z">
      <planning>
        <g2contentType>application/ni+tf+xml</g2contentType>
        <itemClass qcode="ni nat: text" />
        <scheduled>2010-12-25T12:30:00-05:00</scheduled>
        <headline>Robinson: Must preserve bailout funds</headline>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2009-01-26T15:19:11Z">
      <planning>
        <g2contentType>image/jpeg</g2contentType>
        <itemClass qcode="ni nat: picture"></itemClass>
        <scheduled>2010-12-25T12:00:00-05:00</scheduled>
        <edNote>Picture will be Robinson at today's news conference</edNote>
      </planning>
    </newsCoverage>
  </newsCoverageSet>
</planningItem>
```

[<return to Code Listings index](#)

15.6 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.

15.7 Delivered Items - <deliveredItemRef>

A set of properties to identify, locate and describe the delivered Item.

15.7.1 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: href="http://example.com/2008-12-20/pictures/foo.jpg"
@residref	String	The provider's identifier for the target resource, i.e. the @guid of the Item: residref="tag:example.com,2008:PIX:F0020081220098658"
@version	String	The version of the target resource: the @version of the Item
@contenttype	String	The MIME type of the target resource: contenttype="image/jpeg"
@format	QCode	A refinement of the MIME type, taken from a controlled vocabulary
@size	String	Indicates the size of the target resource in bytes: size="3764418"

LISTING 22 Planning Item with delivery at CCL

```
<?xml version="1.0" encoding="UTF-8"?>
<planningItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd"
  guid="urn:newsml:iptc.org:20101025:gbrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="core"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="plinat:newscoverage" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T13:05:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2010-10-16T12:15:00Z</contentCreated>
    <contentModified>2010-10-16T13:35:00Z</contentModified>
  </contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2008-01-26T13:19:11Z">
      <planning>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2009-01-26T15:19:11Z">
      <planning>
        <edNote>Picture scheduled 2010-12-25T12:00:00-05:00</edNote>
      </planning>
      <delivery>
        <deliveredItemRef>
          <title>Henry Robinson pictured today in New York</title>
        </deliveredItemRef>
      </delivery>
    </newsCoverage>
  </newsCoverageSet>
</planningItem>
```



```

        </newsCoverage>
    </newsCoverageSet>
</planningItem>

```

[<<return to Code Listings index](#)

15.7.2 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

15.7.2.1 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 23 Planning Item with delivery at PCL

```

<?xml version="1.0" encoding="UTF-8"?>
<planningItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2_9-spec-All-Power.xsd"
  guid="urn:newsml:iptc.org:20101025:gbmrmdreis4711"
  version="1"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="plinat:newscoverage" />
    <provider literal="IPTC" />
    <versionCreated>2010-10-18T13:05:00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2010-10-16T12:15:00Z</contentCreated>
    <contentModified>2010-10-16T13:35:00Z</contentModified>

```



```
</contentMeta>
  <newsCoverageSet>
    <newsCoverage id="ID_1234568" modified="2008-01-26T13:19:11Z">
      <planning>
        <g2contentType>application/ni tf+xml</g2contentType>
        <itemClass qcode="ni nat: text"/>
        <scheduled>2010-12-25T12:30:00-05:00</scheduled>
        <headline>Robinson: Must preserve bailout funds</headline>
        <edNote>Text 250 words</edNote>
      </planning>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2009-01-26T15:19:11Z">
      <planning>
        <g2contentType>image/jpeg</g2contentType>
        <itemClass qcode="ni nat: picture"></itemClass>
        <scheduled>2010-12-25T12:00:00-05:00</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>
```

[<<return to Code Listings index](#)



16 SportsML-G2

16.1 Introduction

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.

16.2 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.

16.3 Structure

SportsML-G2 content is conveyed within NewsML-G2 Items, (that is News Items, Package Items, Concept Items, Knowledge Items).

The structure of a SportsML-G2 document conveying a single piece of content, or multiple renditions of the same content, closely matches that of a NewsML-G2 News Item, as shown in **Figure 27**.

The <newsItem>, <itemMeta> and <contentMeta> components are implemented as discussed in **Text**, 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.

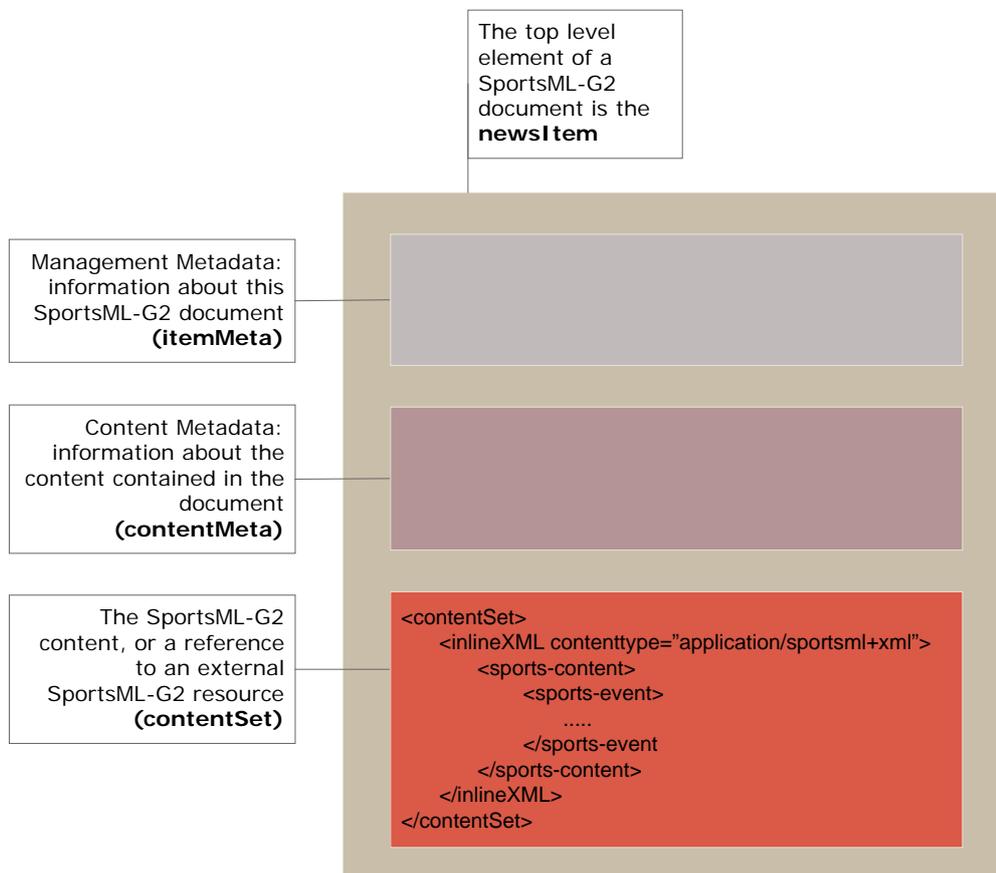


Figure 27: Structure of a SportsML-G2 document matches NewsML-G2

As with NewsML-G2, the <contentSet> component can have child elements of either <inlineXML>, <inlineData>, or <remoteContent>.

The wrapper appropriate to SportsML-G2 is <inlineXML>, which wraps any valid XML content.

16.3.1 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 “2.9”. 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_16.xml" />
<catalog>
  <scheme alias="fixture" uri="http://xmlteam.org/newscodes/fixture#" />
</catalog>
```

16.3.2 Item Metadata

A typical <itemMeta> block²¹ may include the following:

```
<itemMeta>
  <itemClass qcode="innat: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> ← recommended filename
</itemMeta>                                     for the item
```

16.3.3 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 Subject NewsCodes)

```
<contentMeta>
  <contentCreated>
    2009-02-02T11:17:00-05:00 ← When created
  </contentCreated>
  <located qcode="city:Philadelphia"> ← where created
    <broader qcode="reg:Pennsylvania" />
    <broader qcode="cntry:USA" />
  </located>
  <creator qcode="web:sportsnetwork.com"> ← created by
    <name>The Sports Network</name>
  </creator>
  <altId type="idtype:tsn-id" ← alternative ids
    id="sportsnetwork.com-5932656" /> with @type
  <altId type="idtype:revision-id"
    id="1.ml b.com-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
  <genre ← content genre
    type="spcpnat:spfist"
    qcode="spfist: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="xst-genre:tsn-fixture"
    qcode="tsn-fixture:mlbprevi ewxml" />
  <language tag="en-US" />
  <subject qcode="subj:1500000"> ← about the content
    <name xml:lang="en-US">sport</name> (subject codes)
  </subject>
  <subject qcode="subj:15007000">
    <name xml:lang="en-US">Baseball</name>
```

²¹ 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.



```

    <broader qcode="subj:15000000" />
</subject>
<subject qcode="league:l.ml.b.com">
  <name xml:lang="en-US">Major League Baseball</name>
  <broader qcode="subj:15007000" />
  <sameAs qcode="subj:15007001" />
</subject>
<subject type="spcnat:conf"
  qcode="conf:l.ml.b.com-c.national">
  <name xml:lang="en-US">National</name> ← conference (league)
  <broader qcode="league:l.ml.b.com" />
</subject>
<subject type="cpnat:event"
  qcode="event:l.ml.b.com-2007-e.19358" /> ← reference to entry
  in the fixtures CV
<subject type="spcnat:team"
  qcode="team:l.ml.b.com-t.19">
  <name>Philadelphia Phillies</name> ← teams involved
</subject>
<subject type="spcnat:team"
  qcode="team:l.ml.b.com-t.26">
  <name>Arizona Diamondbacks</name>
</subject>
<subject type="cpnat:person"
  qcode="person:l.ml.b.com-p.456">
  <name role="nrol:full">Doug Davis</name> ← players mentioned
  <sameAs qcode="fssID:45679" /> ← "same as" player id
  in another scheme
</subject>
<subject type="cpnat:person"
  qcode="person:l.ml.b.com-p.123">
  <name role="nrol:full">Freddy Garcia</name>
  <sameAs qcode="fssID:45680" />
</subject>
<headline>
  Pitcher Preview: Arizona Diamondbacks (29-23) at ← headline
  Philadelphia Phillies (26-24), 7:05 p.m.
</headline>
<slugline separator="-"
  >AAV! PREVIEW-ARI-PHI ← slugline
</slugline>
</contentMeta>

```

16.3.4 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/
        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



16.3.4.1 Schedules or Fixtures <schedule>

A series of games, usually grouped by date. Also known as a fixture list.

16.3.4.2 Sports Event <sports-event>

Contains metadata and data about a sporting competition – a sporting content that generally ends with a winner

16.3.4.3 Standing <standing>

A series of team or individual records; facts about a team or player's performances during a match, competition, season or career,

16.3.4.4 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.

16.3.4.5 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

16.3.5 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.

16.3.5.1 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="Centigrade" />
  </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.



16.3.5.2 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.

16.3.5.3 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:l.ml.b.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:l.ml.b.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>
```

16.3.5.4 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:l.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-goals-scored="0" drop-goals-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>
```

16.3.5.5 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" />
```

16.3.5.6 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>
<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>
```

16.3.5.7 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

16.3.5.8 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:fi fa02345"
      position="referee">
      <name>Keith Burkinshaw</name>
    </official-metadata>
    <official-stats>
      <rating rating-issuer="FIFA" rating-maximum="8"
        rating-value="8" />
    </official-stats>
  </official>
</officials>
```

16.3.5.9 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 24 A complete sample SportsML-G2 document

```
<?xml version="1.0" encoding="UTF-8"?>
<newsItem guid="tag:xml.team.com,2009:xt.5932656-pitcher-preview"
  version="3" xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
  XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  xmlns:xts="www.xml.team.com" standard="NewsML-G2" standardversion="2.9"
  conformance="power" xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef
    href="http://www.xml.team.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="inat:text" />
    <provider qcode="web:www.xml.team.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">
      <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-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
    <genre type="spcnat:spfxt" qcode="spfxt:pitcher-preview">
      <name xml:lang="en-US">Game Pitcher Preview</name>
      <broader qcode="dclass:event-summary" />
    </genre>
    <genre type="spcnat:dclass" qcode="dclass:event-summary" />
    <genre type="xts-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewxml" />
    <language tag="en-US" />
    <subject qcode="subj:1500000">
      <name xml:lang="en-US">sport</name>
```



```

</subject>
<subject qcode="subj:15007000">
  <name xml:lang="en-US">Baseball</name>
  <broader qcode="subj:15000000" />
</subject>
<subject qcode="league:1.ml.b.com">
  <name xml:lang="en-US">Major League Baseball</name>
  <broader qcode="subj:15007000" />
  <sameAs qcode="subj:15007001" />
</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">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.ml.b.com-2009-e.19358"
          date-coverage-value="1.ml.b.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.ml.b.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.ml.b.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.ml.b.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.ml.b.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>

```

[<<return to Code Listings index](#)

16.3.6 SportsML-G2 Packages

To convey a package of SportsML-G2 items, simply use the NewsML-G2 <packageItem>, as covered in **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 (NITF²²) document, conveyed in a NewsML-G2 News Item, as shown in **LISTING 25**.

LISTING 25 Sports story in NewsML-G2/NITF

```

<?xml version="1.0" encoding="UTF-8"?>
<newsItem
  guid="tag:xmlteam.com,2009:xt.5932656-preview"
  version="3"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  xmlns:xts="www.xmlteam.com"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="inat:text" />
    <provider qcode="web:www.xmlteam.com" />
    <versionCreated>2010-10-19T11:45:00-04:00</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: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" />
    <language tag="en-US" />
    <subject qcode="subj:15000000">
      <name xml:lang="en-US">sport</name>
    </subject>

```

²² The IPTC's XML standard for marking up text.



```
<subject qcode="subj:15007000">
  <name xml:lang="en-US">Baseball</name>
  <broader qcode="subj:15000000" />
</subject>
<subject qcode="league:1.mlb.com">
  <name xml:lang="en-US">Major League Baseball</name>
  <broader 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>
  <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>Doug Davis</name>
  <sameAs qcode="fssID:45679" />
</subject>
<subject type="cpnat:person" qcode="person:1.mlb.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/ni tf+xml">
    <ni tf xmlns="http://iptc.org/std/NITF/2006-10-18/"
      xsi:schemaLocation="http://iptc.org/std/NITF/2006-10-18/
      ../XSD/NITF/3.5/specification/ni tf-3-5.xsd">
      <body>
        <body.head>
          <headline>
            <hl 1>Arizona Diamondbacks (29-23) at Philadelphia Phillies (26-24),
              7:05p.m.</hl 1>
          </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>
      </ni tf>
    </inlineXML>
  </contentSet>
</contentMeta>
```



```
</body.content>
</body>
</niftf>
</inlineXML>
</contentSet>
</newsItem>
```

[<<return to Code Listings index](#)

16.3.6.1 Referencing the packaged items

Each of the documents listed have 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.

16.3.6.2 Package Item <packageItem>

The root <packageItem> element is coded as shown (some line-breaks inserted for ease of reading) and is similar to all other Items. Appropriate changes are noted below:

```
<?xml version="1.0" encoding="UTF-8"?>
<packageItem guid="urn:iptc.org:20060101:sports1" version="1"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/ ← schema change
  XSD/NewsML-G2_2.9-spec-All-Power.xsd "
  xmlns:xts="www.xmlteam.com"
  conformance="power"
  standard="NewsML-G2" ← standard and
  standardversion="2.9" standard version
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
```

16.3.6.3 Item Metadata and Content Metadata

The <itemMeta> and <contentMeta> components are similar to the previous examples. Some of the Content Metadata (located, creator, and subject) is common to both items and is repeated in the <contentMeta> of the package as an aid to processing by the receiver.

16.3.6.4 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 property 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 28**

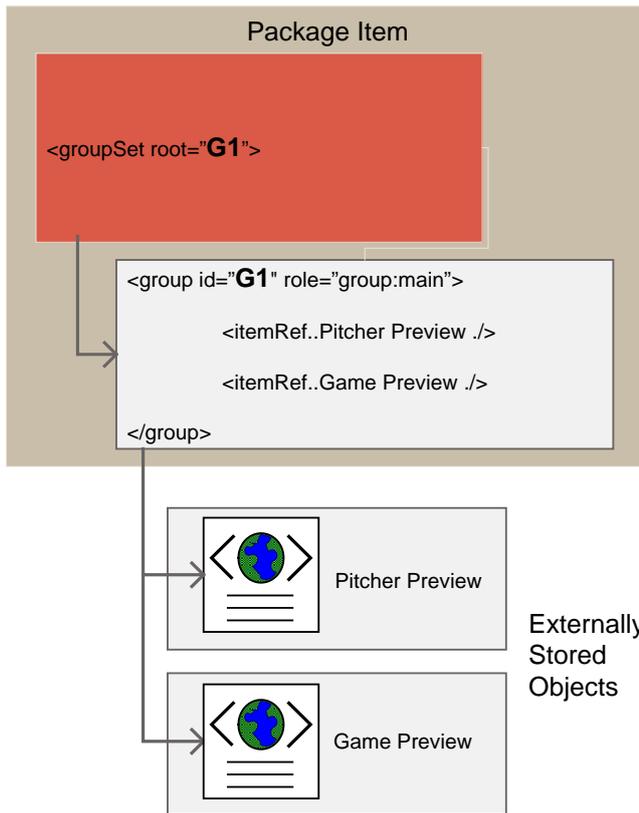


Figure 28: Diagram of Package Group structure

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: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/ni tf+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>

```



The complete Package Listing is shown below.

LISTING 26 SportsML-G2 Package

```

<?xml version="1.0" encoding="UTF-8"?>
<packageItem guid="urn:iptc.org:20060101:sports1" version="3"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  xmlns:xts="www.xmlteam.com"
  conformance="power"
  standard="NewsML-G2"
  standardversion="2.9"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="inat: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">
      <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-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>
      <broader qcode="subj:15000000" />
    </subject>
    <subject qcode="league:1.mlb.com">
      <name xml:lang="en-US">Major League Baseball</name>
      <broader 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>
      <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>
  </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:mlbpreviexml" />
        <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="xtns-genre:tsn-fixture" qcode="tsn-fixture:mlbpreviewsxml" />
<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>
```

[<<return to Code Listings index](#)



This page intentionally blank

17 Exchanging News: News Messages

17.1 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.

17.2 Structure

A diagram of the structure of a News Message and its payload of Items is shown in **Figure 29** below.

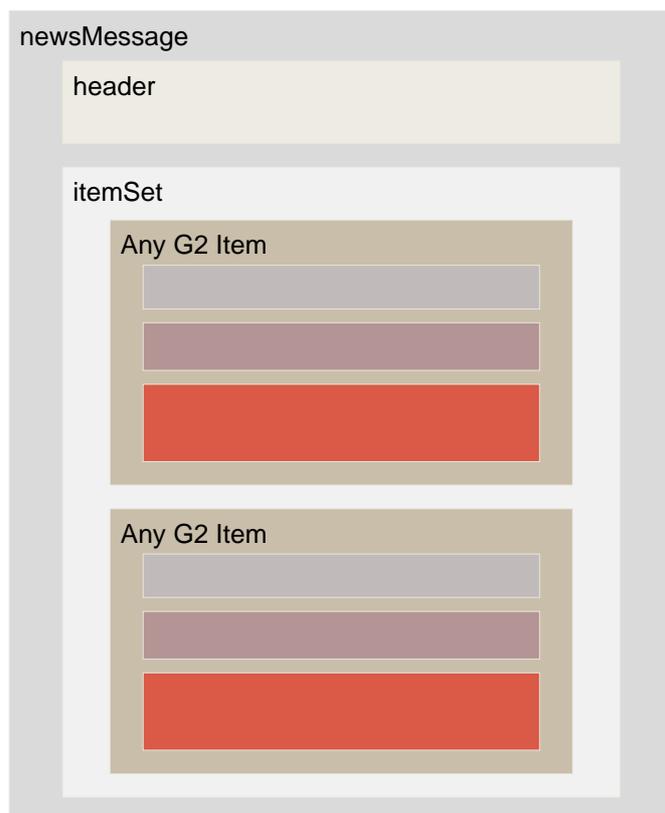


Figure 29: News Messages can convey any number of any kind of G2 Item

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.



17.2.1 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 QCode or literal values, and also use other QCode type properties to add further detailed information, e.g.:

```
<sender qcode="s ndr: rtr">thomsonreuters.com</sender>
```

or

```
<sender literal="thomsonreuters.com">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 7.3.2 for details on implementing Catalogs)²³. 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.
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.

17.2.1.1 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>2009-02-01T11:17:00.000Z</sent>
```

²³ <catalog> or <catalogRef> are optional for News Messages, but mandatory for Items.



17.2.1.2 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="nmsg" uri="http://cv.example.com/newscodes/newsmgsignal/" />  
</catalog>
```

17.2.1.3 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/catalog.IPTC-G2-Standards_16.xml" />
```

17.2.1.4 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>
```

or

```
<sender role="nmrole:aggregator type="ci:org"  
  qcode="nmsdr:rtr">thomsonreuters.com</sender>
```

17.2.1.5 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.

```
<transmitID>87e3841c-430e-4da5-bed6-28cae4e8fe3b</transmitID>
```

17.2.1.6 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.

```
<priority>4</priority>
```

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.

17.2.1.7 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.

```
<origin>MMS_3</origin>
```

or

```
<origin qcode="nmorgn:mms3" />
```

17.2.1.8 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" />
```

17.2.1.9 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">foo.bar@example.com</destination>
```

The diagram shown below in **Figure 30** illustrates the application of Destination and Channel in a syndication workflow.

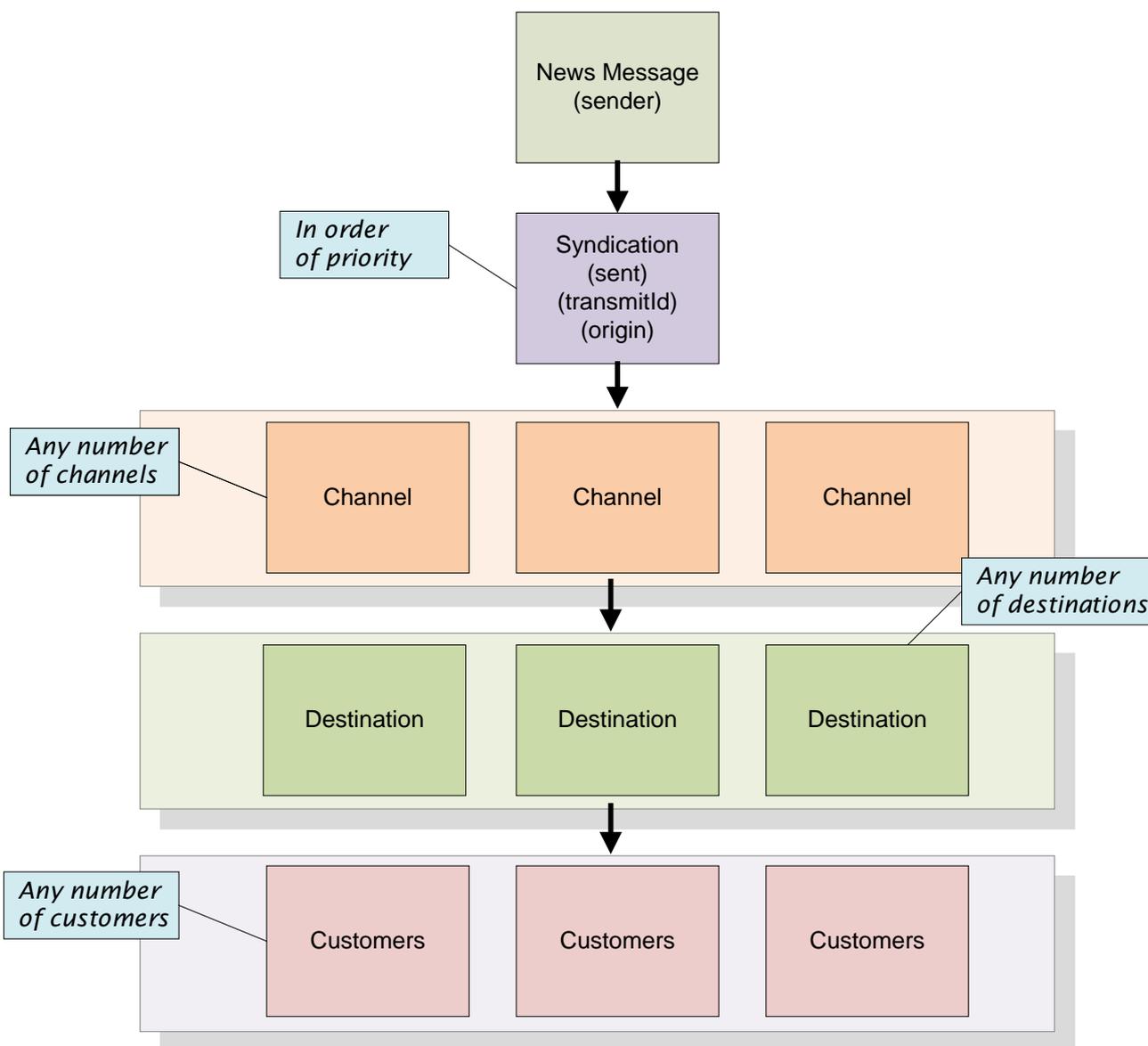


Figure 30: Syndication workflow using Destination and Channel

Note that there are many possible paths through to each customer: a News Message could be delivered to on many Channels via many Destinations and then to many subscribing Customers.



17.2.1.10 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">2009-02-01T11:17:00.000Z</timestamp>
<timestamp role="transmitted">2009-02-01T11:17:00.100Z</timestamp>
```

17.2.1.11 Signal <signal>

Optional, repeatable; at PCL only, the <signal> property can be used to indicate any special handling instruction to the receiver's G2 processor:

```
<signal qcode="nmsg:atomic" />
```

There is a recommended IPTC News Message Signal NewsCode CV with a Scheme URI of:
<http://cv.iptc.org/newscodes/newsmgsignal/>

The recommended scheme alias is "nmsg" 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.

17.2.2 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 27 News Message conveying a complete News Package

```
<?xml version="1.0" encoding="UTF-8"?>
<newsMessage xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-NewsMessage-Power.xsd">
  <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_16.xml" />
    <sender>thomsonreuters.com</sender>
    <transmissionId>tag:reuters.com,2009:newsml_OVE488500-PKG</transmissionId>
    <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="nmsg:atomic" />
  </header>
  <itemSet>
    <packageItem>
      <itemRef residref="N1" />
      <itemRef residref="N2" />
      <itemRef residref="N3" />
      <itemRef residref="N4" />
    </packageItem>
  </itemSet>
</newsMessage>
```



```
</packageItem>
<newsItem guid="N1"> . . . . . </newsItem>
<newsItem guid="N2"> . . . . . </newsItem>
<newsItem guid="N3"> . . . . . </newsItem>
<newsItem guid="N4"> . . . . . </newsItem>
</itemSet>
</newsMessage>
```

[<<return to Code Listings index](#)



18 Migrating IPTC 7901 to NewsML-G2

18.1 Introduction

The text transmission standard IPTC 7901 (and ANPA1312, with which it is closely associated) have been a mainstay of the exchange of text for 30 years and is still heavily used. 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.

18.1.1 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 <destination> property of News Message:</p> <pre><destination>FRS</destination></pre> <p>An alternative property would be <channel></p> <pre><channel>FRS</channel></pre> <p>Within a News Item, there is a <service> 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 <transmitId> 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> in the Administrative Metadata group of <contentMeta></pre> <pre><urgency>3</urgency></pre> <pre><priority> in the <header> of News Message.</pre> <pre><priority>3</priority></pre> <p>If the 7901 field is mapped from an editorial system and is under the control of journalists, then <urgency> 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</p>



Field Name	Example	NewsML-G2 Property and Example
		<p>categories for NewsML-G2 would create a Controlled Vocabulary mapping 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: P0L"> <name>Pol i t i c s</name> </subject></pre>
Word Count	195	<p>The maximum value allowed in IPTC 7901 is 9999. Word Count is one of the properties of News Content Characteristics in NewsML-G2 and for the plain text of IPTC 7901 is an attribute of the Inline Data component <inlineData></p> <pre><i n l i n e D a t a wordcount=" 195" > Text text text </i n l i n e D a t a ></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> <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><sl u g l i n e separator="- " >US- P O L I T I C S / B U S H </sl u g l i n e ></pre>
Keyword/Catch-Line		

18.1.2 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.

18.1.3 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:

```
<versi onCreated>2009- 02- 09T12: 30: 00Z</versi onCreated>
```



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. This allows just the date to be expressed, with optional time and time zone.

| <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



19 News Industry Text Format (NITF)

19.1 Introduction

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 1.x and NewsML-G2:

- ❖ NITF provides a mark-up structure for text; NewsML 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 this Implementation Guide Chapters [Anatomy of NewsML-G2](#) and [Text](#).

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.



19.2 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.

19.2.1 Root element <nitf>

The root element may uniquely identify the document using @uno, which is equivalent to NewsML-G2 @guid property

19.2.2 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.
<tobject>	<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 19.2.2.1 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

19.2.2.1 <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> 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



NITF element	NewsML-G2 equivalent	Comment
		<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 >	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.
<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.



NITF element	NewsML-G2 equivalent	Comment
		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 >	<subject>	NewsML-G2 has no specific equivalent container for keywords. Providers may use a @literal value of <subject> to express a keyword.
< 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>

19.3 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 <h1> and subhead <h2>
<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 will need to consult with their provider, either directly or via documentation.</p>
<rights >	<rightsInfo>	See 19.2.2.1 above



NITF element	NewsML-G2 equivalent	Comment
<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 19.2.2.1 above

19.4 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 28 An NITF marked-up article conveyed in <inlineXML>

```
<contentSet>
  <inlineXML content type="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>
```

[<<return to Code Listings index](#)

19.5 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> signifies 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.



This page intentionally blank

20 Mapping Embedded Photo Metadata to NewsML-G2

20.1 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.metadataworkinggroup.org.

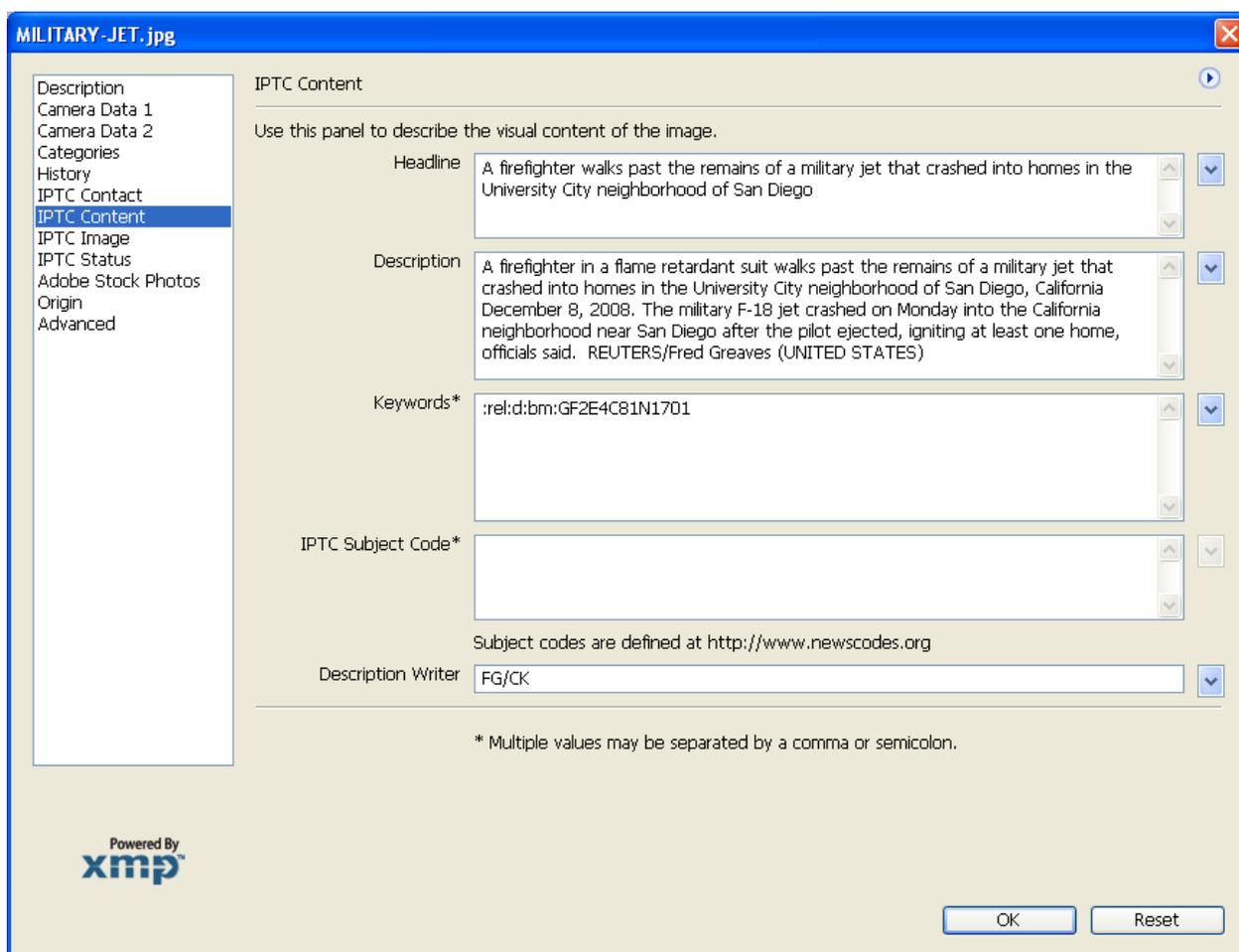


Figure 31: One of the IPTC Core metadata panels from an image opened in Adobe Photoshop.



20.2 IPTC Metadata for XMP

To use IPTC metadata in Adobe Photoshop CS (aka 8.0) and onwards, a set of custom panels must be downloaded from the IPTC Web site and installed into the application. Links to the panels and to the documentation are located at http://www.iptc.org/site/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.

20.3 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.

20.4 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.

20.5 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.

20.6 IIM resources

The IPTC web site has an IIM Home page (www.iptc.org/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.

20.7 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, which can be found at:

<http://www.iptc.org/std/photometadata/specification/>

20.8 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. See the IIM Specification <http://www.iptc.org/std/IIM/4.1/specification/IIMV4.1.pdf> for details of the record structure)

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 [20.9](#)

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	itemMeta/altId	If the reference is a Job



DataSet	IIM Name	IPTC Core Name (XMP)	NewsML-G2 Property	Note
	Transmission Reference	Reference		ID use itemMeta/memberOf
2:105	Headline	Headline	contentMeta/headline	
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

20.9 IIM to NewsML-G2 Mapping Example

The illustration below shows the IPTC Content Panel, one of the IPTC Core custom panels in Adobe Photoshop's File Info dialog. 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.

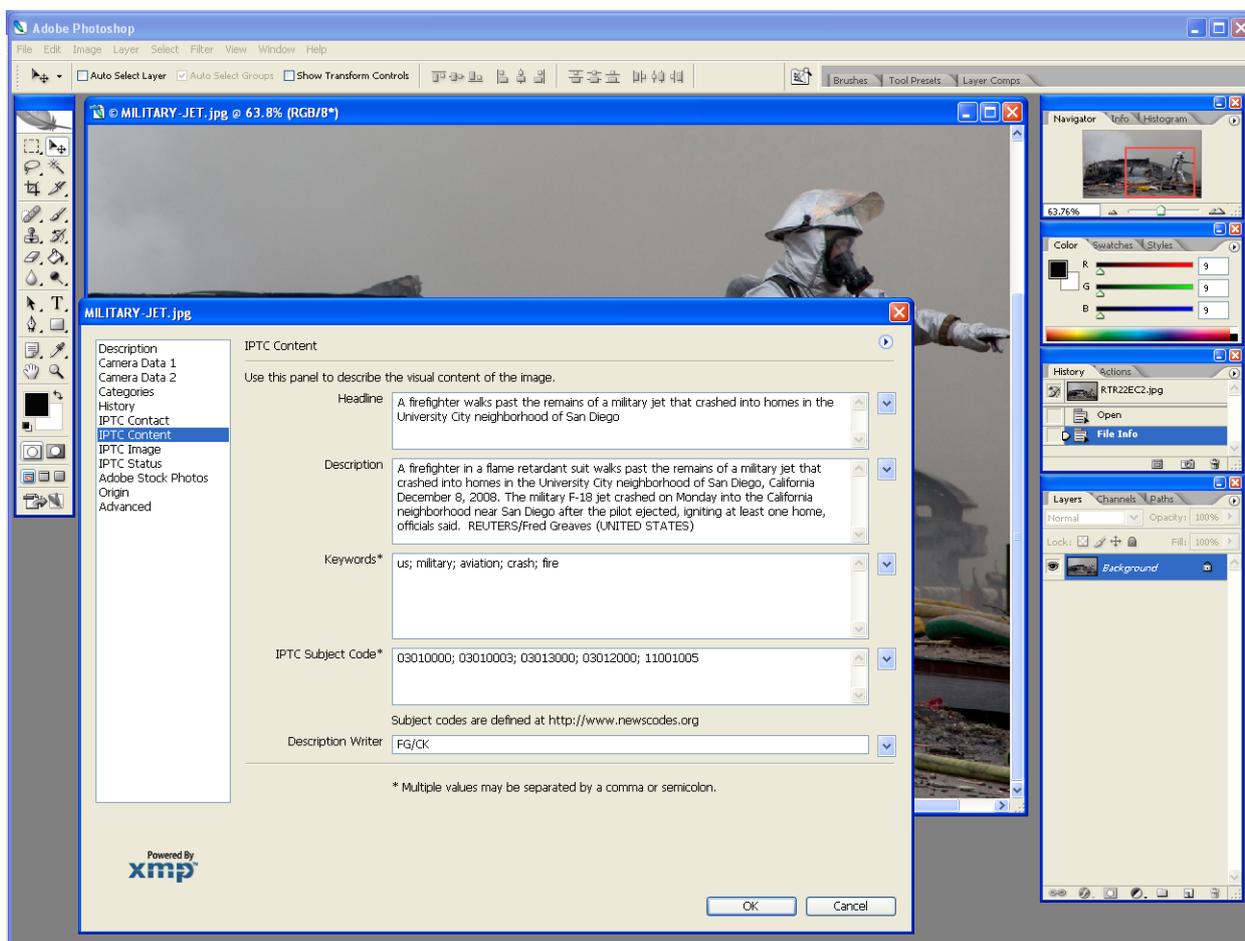


Figure 32: Opening File Info panel of the image (photo and metadata ©Thomson Reuters)

DataSet	IIM Name	Value	Ref
2:05	Object Name	USA-CRASH/MILITARY	20.9.1
2:10	Urgency	4	20.9.2
2:15	Category	A	20.9.3
2:20	Supplemental Category	DEF DIS	20.9.3
2:25	Keywords	us	20.9.3
2:25	Keywords	military	



DataSet	IIM Name	Value	Ref
2:25	Keywords	aviation	
2:25	Keywords	crash	
2:25	Keywords	fire	
2:40	Special Instruction	NO ARCHIVE OR WIRELESS USE	20.9.5
2:55	Date Created	20081208	20.9.6
2:60	Time Created	133000-0800	20.9.6
2:80	By-Line	John Smith	20.9.7
2:85	By-line Title	Staff Photographer	20.9.7
2:90	City	San Diego	20.9.8
2:92	Sub-location	University City	20.9.8
2:95	Province/State	CA	20.9.8
2:100	Country/Primary Location Code	USA	20.9.8
2:101	Country/Primary Location Name	United States	20.9.8
2:103	Original Transmission Reference	SAD02	20.9.9
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	20.9.10
2:110	Credit	Acme/John Smith	20.9.7
2:115	Source	Acme News LLC	20.9.7
2:116	Copyright Notice	© Copyright Acme News LLC. Click For Restrictions - http://www.acmenews.com/legal	20.9.11
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.	20.9.12
2:122	Writer/Editor	CK	20.9.13

20.9.1 Object Name / Title

The <title> property 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="inat:picture" />
  <provider literal="reuters.com" />
  <versionCreated>
    2008-12-09T02:20:00Z
  </versionCreated>
  <title>USA-CRASH/MILITARY</title> ← Title
</itemMeta>
```

Some providers may additionally map this DataSet to the NewsML-G2 <slugline> property in <contentMeta>:

```
<slugline>USA-CRASH/MILITARY</slugline>
```



20.9.2 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>
```

20.9.3 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.

20.9.4 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 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 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 the <keyword> property.

```
<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.



20.9.5 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 literal="Acme News LLC" />
  <copyrightNotice>(c) Copyright Acme News LLC 2009. For conditions of use see
    http://www.acmenews.com/legal
  </copyrightNotice>
  <usageTerms xml:lang="en">NO ARCHIVAL OR WIRELESS USE</usageTerms>
</rightsInfo>
```

20.9.6 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).²⁴

```
<contentMeta>
  <contentCreated>2008-02-08T13:30:00-08:00</contentCreated>
  ...
</contentMeta>
```

20.9.7 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)

```
<contentMeta>
  <contentCreated>2008-02-09T13:30:00-08:00</contentCreated>
  <creator role="crol:photog" qcode="pers:JS001">
    <name>John Smith</name>
    <related qcode="fcode:jobtitle">
      <name>Staff Photographer</name>
    </related>
  </creator>
```

²⁴ If present, the time part must be used in full, with time zone, and ONLY in the presence of the full date.



```
</creator>
...
</contentMeta>
```

Credit should be mapped to the <creditline> child property of <contentMeta>. There is a <provider> property of <itemMeta>, but the Credit does not 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 Smith</creditline>
```

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>
...
</rightsInfo>
```

20.9.8 City, Province/State and Country

As discussed in **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 would be expressed using a series of <subject> properties as follows:

```
<subject type="cpnatexp:geoAreaSublocation" literal="int001">
  <name>University City</name>
  <broader type="cpnatexp:geoAreaCity" literal="int002" />
</subject>
<subject type="cpnatexp:geoAreaCity" literal="int002">
  <name>San Diego</name>
  <broader type="cpnatexp:geoAreaStateProv" literal="int003" />
</subject>
<subject type="cpnatexp:geoAreaStateProv" literal="int003">
  <name>CA</name>
  <broader type="cpnatexp:geoAreaCountry" literal="USA" />
</subject>
```



```
<subject type="cpnatexp: geoAreaCountry" literal="USA" >
  <name>United States</name>
</subject>
```

The properties are 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.

Use of <broader> is preferred to <narrower> in this case because it would be easier to add a new child into the hierarchy without altering the existing code.

An alternative to this method would be to use a single <subject> property for the location, linked to an <assert> block using a @literal identifier:

```
<contentMeta>
  . . . .
  <subject type="cpnatexp: geoAreaSublocati on" literal="int001" /> ← @literal
  . . . .
</contentMeta>
<assert literal="int001"> ← @literal
  <type qcode="cpnatexp: geoAreaSublocati on" />
  <name>Uni versi ty Ci ty</name>
  <broader literal="int002" type="cpnatexp: geoAreaCi ty">
    <name>San Di ego</name>
  </broader>
  <broader literal="int003" type="cpnatexp: geoAreaProvState">
    <name>Cal i forni a</name>
  </broader>
  <broader literal="USA"25 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 a NewsML-G2 document. 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.

20.9.9 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 properties: @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: i i m">SAD02</al tId>
```

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: j obref" literal="SAD02" />
```

²⁵ If using a @literal identifier for a country, the Country Code, if available, should be used as the identifier. The use of QCode identifiers is to be preferred if possible and practical.

The IPTC recognizes that some well-known CVs, such as those maintained by the International Standards Organization (ISO) are widely used in news exchange. Rather than duplicate this work, the IPTC catalog contains references to these CVs, including the 2-letter and 3 letter country codes. The recommended aliases for these schemes are iso3166-1a2 and iso3166-1a3 respectively. For more information, see the page <http://cvx.iptc.org> hosted by the IPTC. In the full code listing at the end of this Chapter, the Country is identified by a QCode.



20.9.10 **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 the @xml:lang property to declare the language and variant "en-US".

20.9.11 **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>
```

20.9.12 **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 flame
retardant walks past the remains of a military jet that crashed into homes
suit in the University City neighborhood of San Diego, California December 8,
2008. The F-18 jet crashed on Monday into the California neighborhood military
near San Diego after the pilot ejected, igniting at least one home,
officials said.
</description>
```

20.9.13 **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 29 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"?>
<newsItem
  guid="tag:acmenews.com,2008:WORLD-NEWS:USA20090209098658"
  version="3"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
  XSD/NewsML-G2_2.9-spec-All-Power.xsd"
  standard="NewsML-G2"
  standardversion="2.9"
  conformance="power">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.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="inat:picture" />
```



```

<provider literal="acmenews.com" />
<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="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>
<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>
<broadener qcode="iso3166-1a3:USA" type="cpnatexp:geoAreaCountry">
<name>United States</name>
</broadener>
</assert>
<contentSet>
<remoteContent
href="http://www.acmenews.com/content/pictures/hires/20090209/USA-CRASH-
MILITARY_001_HI.jpg"
rendition="rnd:highRes" size="3764418" contentType="image/jpeg" width="2500"
height="1445" colourSpace="colsp:USSWOPv2" />
</contentSet>
</newsItem>
    
```

[<return to Code Listings index](#)

20.10 Reconciling NewsML-G2 with embedded metadata

Guidance was discussed in [Reconciling photo metadata](#) for situations 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.



21 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. A news content specific Profile for Open Digital Rights Language (ODRL) version 2.0 has been defined, 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. This ODRL profile enables a wide range of usage policies to be expressed in NewsML-G2. See the IPTC web site www.iptc.org for further details on that standard, named RightsML.

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 [21.5](#) below.

21.1 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 literal="RTRS" />
  <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>
```

This Core (CCL) implementation of `<rightsInfo>` remains unchanged in the latest version of NewsML-G2.

21.2 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 keywording (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.

21.3 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”.

21.3.1 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) <contentSet> in a News Item, b) <concept> in a Concept Item, c) <conceptSet> in a Knowledge Item and d) <groupSet> (except all children of <itemRef>) in a Package Item. Any parts of content which are described by <partMeta> 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 <itemRef> elements in a Package Item.

21.3.2 Rights Aspect NewsCodes

Scheme URI: <http://cv.iptc.org/newscodes/raspect/>

Recommended scheme alias: raspect

Code	Definition
values	The <rightsInfo> element makes an assertion about metadata values including the details of concepts. This aspect applies only to the Rights Info Scope of “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”.



21.4 Use cases for the extended <rightsInfo> model

21.4.1 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="ri scope: content" >
  <copyrightHolder literal="EPIX">
    <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="ri scope: metadata" >
  <copyrightHolder literal="PLIB">
    <name>The Picture Library</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright The Picture Library Ltd 2008.
  </copyrightNotice>
</rightsInfo>
```

21.4.2 Separating rights for part of the metadata

As in **21.4.1** 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 **21.3.1**).
[copyrightHolder!]

```
<newsItem>
...
<!-- All: Third Party -->
<rightsInfo>
  <copyrightHolder literal="EPIX">
    <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 literal="PLIB">
    <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>
...
```



21.4.3 Separating rights for aspects of part of the metadata

As in [21.4.2](#) 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 literal="EPIX">
    <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="respect:values"> [values, not codes?]
  <copyrightHolder literal="IPTC">
    <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="respect:selection">
  <copyrightHolder literal="PLIB">
    <name>The Picture Library</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright The Picture Library Ltd 2008.
  </copyrightNotice>
</rightsInfo>
...
<contentMeta>
...
  <subject id="id001" type"cpnat:abstract" qcode="subj:01027000" />
  <subject id="id002" type"cpnat:abstract" qcode="subj:02001000" />
...
</contentMeta>
...
```

21.4.4 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 literal="AGY">
    <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 literal="RTRS">
    <name>Thomson Reuters</name>
  </copyrightHolder>
  <copyrightNotice>(c) Copyright Thomson Reuters 2010.
  </copyrightNotice>
</rightsInfo>
```



```

...
</contentMeta>
<partMeta parti d="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>

```

21.4.5 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="ri scope: content" aspect="raspect: structure">
  <copyrightHolder literal="PROV">
    <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="ri scope: content" aspect="raspect: selection">
  <copyrightHolder literal="CONT">
    <name>The News Selector</name>
  </copyrightHolder>
  <copyrightNotice>
    (c) Copyright asserted to the selection of the <itemRef>s.
  </copyrightNotice>
  <usageTerms xml:lang="en">
    Usage terms in natural language
  </usageTerms>
</rightsInfo>
<!-- Metadata: News Provider -->
<rightsInfo scope="ri scope: metadata" >
  <copyrightHolder literal="PROV">
    <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>.

21.5 Processing model for extended Rights Info

As these cross-references and additional attributes not only add useful features but also some complexity, the IPTC provides a detailed processing model. Three different variants are provided; each one takes a different view on the relationship of rights-related information and the (metadata and content) particles of a NewsML-G2 Item.

21.5.1 Adding Rights Information expressions: the news provider's view

A news provider is now able to express rights in a more granular way:

- ❖ @scope may be used to express separate rights to the metadata and content of an Item.
- ❖ @aspect may be used to indicate whether the rights expression applies to metadata values, or to the application of those values. It may also be used to indicate whether rights are reserved for the selection of content in an Item (e.g. a Package Item) or to the structure of the content.
- ❖ Using @idrefs, a rights expression can be targeted at individual XML elements, either in the metadata or via <partMeta> to parts of the content (provided the markup supports @id attributes).

21.5.1.1 Implementing Rights Info @scope

The scope of a <rightsInfo> wrapper in general can be expressed in a targeted way. Thus, the Rights Info is about:

- ❖ the *Item as a whole* if a @scope attribute is not present
- ❖ the *metadata properties* if a @scope attribute is added with a value of "riscope:metadata"
- ❖ the *content* if a @scope attribute is added with a value of "riscope:content"

The scope of a <rightsInfo> wrapper can be more specifically targeted by adding @idrefs containing the @ids of the elements to which the rights refer. This specific addressing of elements overrides <rightsInfo> expressions which use the mechanism described above.

If the referenced XML element is a partMeta element then:

- ❖ If a @scope attribute is not present, the <rightsInfo> applies to *both* the *content* described by this <partMeta> element and to the *metadata children* of this <partMeta> element.
- ❖ If a @scope attribute is present its value(s) *determines* whether <rightsInfo> applies to the *content* described by this <partMeta> element or to the *metadata children* of this <partMeta> element.



Use caution when combining @scope with @idrefs. Some elements may be unintentionally "orphaned" because the rights expression will only apply to the element(s) identified by a corresponding @id.

IDs of the following XML elements *may* be included into the list of values of @idrefs:

- ❖ all metadata properties, except the wrapper elements as <contentMeta> and <itemMeta>. (See below)
- ❖ the child elements of a News Item <contentSet>, <inlineXML>, <inlineData> and <remoteContent> (when they provide different renditions of the same content, to which different rights may apply - e.g. the hi-res rendition of a picture may have different rights to the web-resolution rendition).
- ❖ the child <concept> elements of a Knowledge Item's <conceptSet>



- ❖ the child <group> elements of a Package Item's <groupSet>.

IDs of the following XML elements *may NOT* be included into the list of values of @idrefs:

- ❖ all child elements of a News Item's <inlineXML>, as they contain only parts of the content. In this case a <partMeta> element must be used to describe this part and the value of the @partid attribute of this <partMeta> element must be added to the list of values of the @idref attribute of the <rightsInfo> element.
- ❖ the wrapper elements <contentMeta> and <itemMeta>. Any rights expressed for these wrapper elements are NOT inherited by their child elements so they should not be used.

21.5.1.2 Implementing Rights Info @aspect

A specific aspect of rights, as defined by the Aspect NewsCodes CV, can be applied to a <rightsInfo> element:

- ❖ If an @aspect attribute is not present then *all aspects* from the Rights Aspect NewsCodes apply.
- ❖ If an @aspect attribute is present then *only the aspects* from the Rights Aspect NewsCodes listed in the attribute apply.
- ❖ If a target does not support a specific aspect which is listed in the @aspect attribute then this aspect *should be ignored* for this target.

The flowchart on the following page illustrates the rules for adding Rights Information expressions to parts of an Item:



Key:
CON: Content.
 Item Part ::= full CON | one CON rendition | part of CON described by partMeta | single MET property.
MET: Metadata.
RI: rightsInfo element.
RS: Result Set.
TRS: Temporary RS.
A += B : A = A + B.

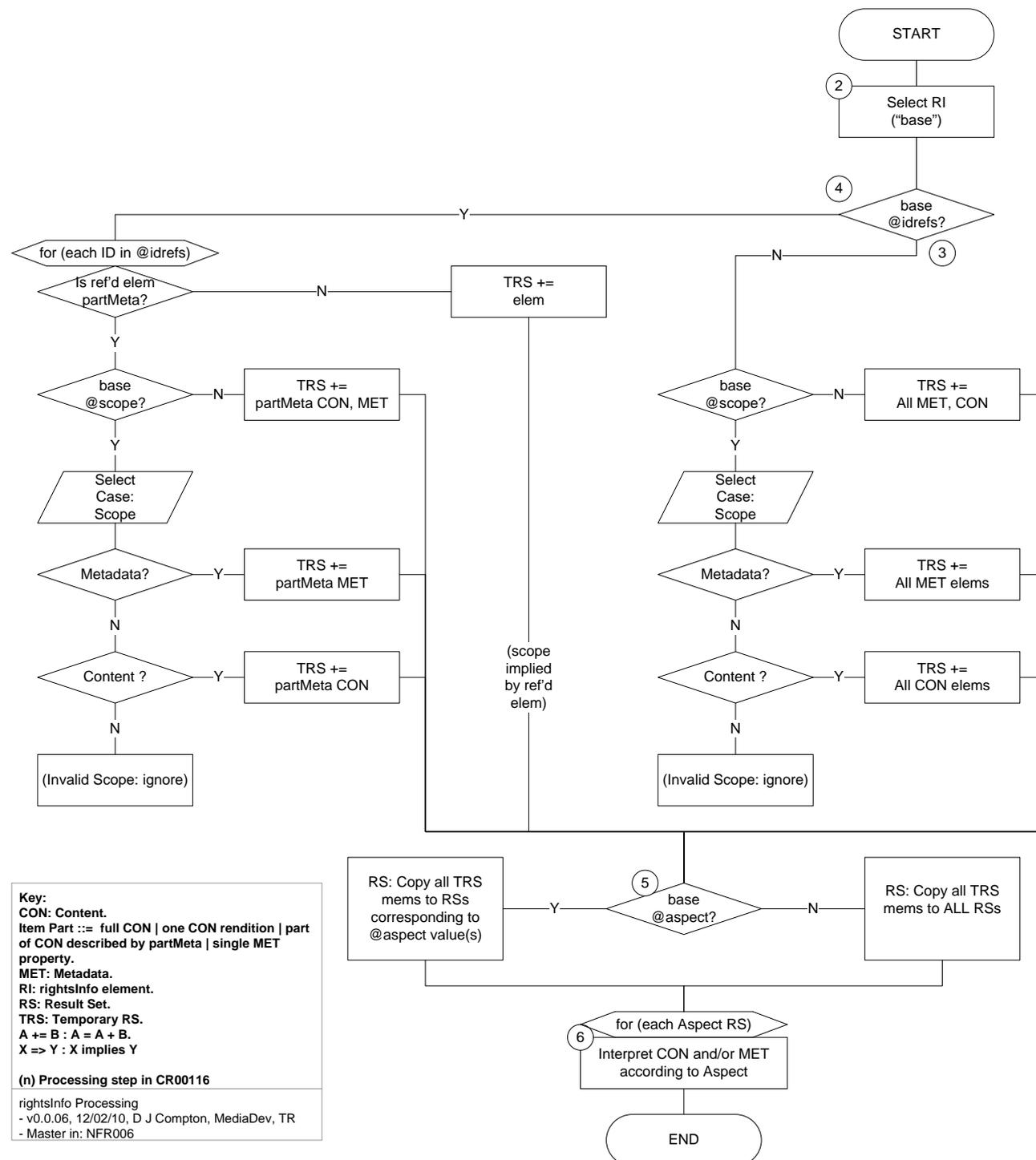
(n) Processing step in CR00116

rightsInfo Processing
 - v0.0.04, 15/09/10, D J Compton, MediaDev, TR
 - Master in: NFR006

Figure 33: How to apply rightsInfo elements referencing only a fraction of a NewsML-G2 Item

21.5.2 Processing Diagram: Rights Holder view

The following flowchart shows how a copyright holder or other interested party can identify the parts of an Item to which a specific <rightsInfo> expression refers.



Key:
CON: Content.
Item Part ::= full CON | one CON rendition | part of CON described by partMeta | single MET property.
MET: Metadata.
RI: rightsInfo element.
RS: Result Set.
TRS: Temporary RS.
A += B : A = A + B.
X => Y : X implies Y

(n) Processing step in CR00116

rightsInfo Processing
 - v0.0.06, 12/02/10, D J Compton, MediaDev, TR
 - Master in: NFR006

Figure 34: Processing diagram to identify the parts to which a rightsInfo refers

21.5.2.1 Rights Holder Flowchart Narrative

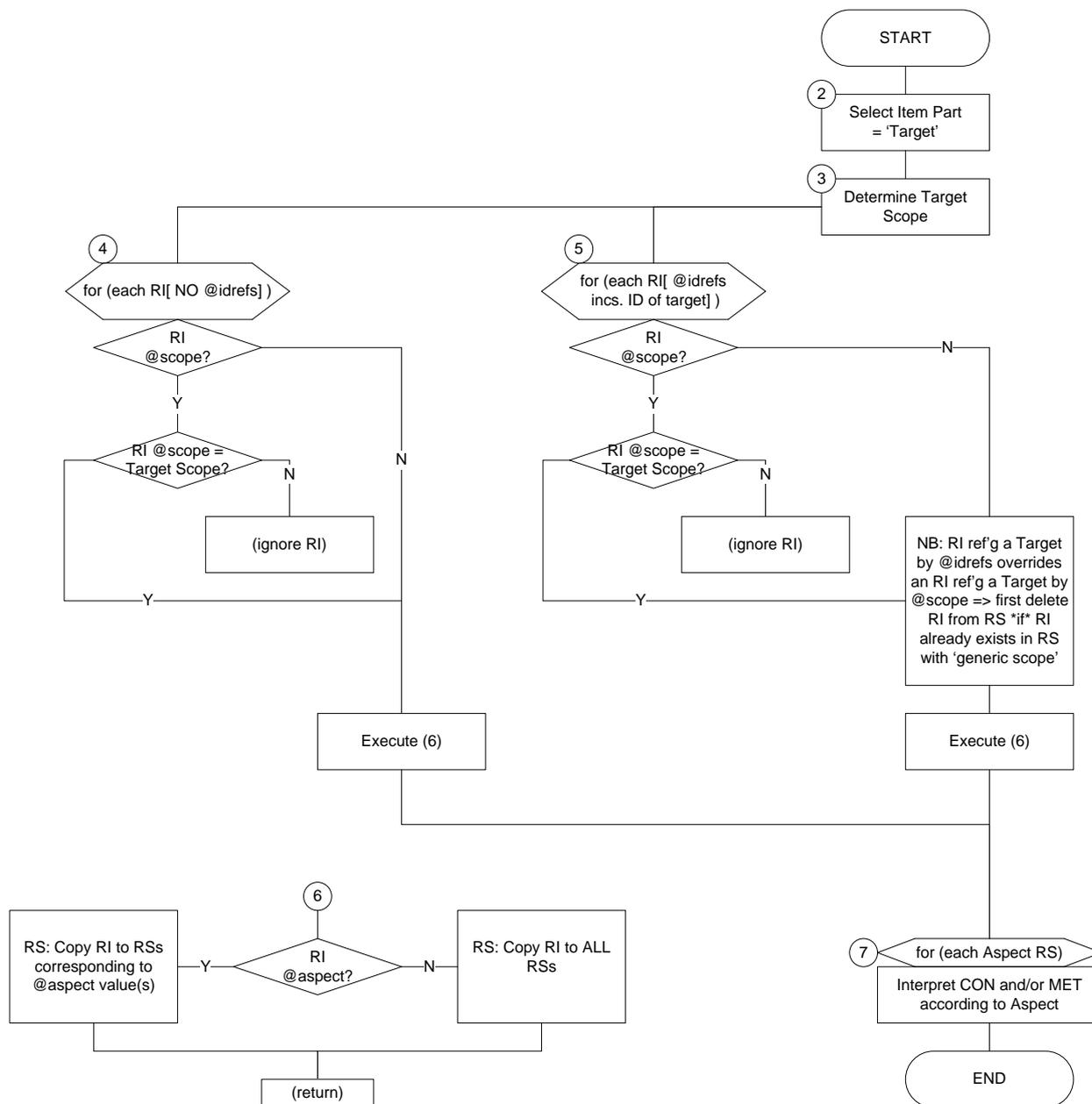
1. The goal of the processing: the result will be multiple sets of elements and/or parts of content which all are governed by a <rightsInfo> expression. Each of the sets a) corresponds to one of the Rights Aspect NewsCodes, and b) may be empty after the processing if no corresponding parts of an item were found.



2. Select the <rightsInfo> element to be processed; this is the “base” for all subsequent processing steps.
3. If no @idrefs attribute exists in the base:
 - 3.1. If a @scope attribute is not present: all the content and all metadata properties of this item are governed by the base’s rights expression; they all should be included into a temporary result set. Continue with step 5.
 - 3.2. If a @scope attribute is present:
 - 3.2.1. If its value is “riscope:metadata”: only metadata properties are in the scope of this <rightsInfo> element, add only all metadata elements of this item to a temporary result set. Continue with step 5.
 - 3.2.2. If its value is “riscope:content”: only content is in the scope of this <rightsInfo> element, add only all content of this item to a temporary result set. Continue with step 5.
4. If an @idrefs attribute is present in the base:
 - 4.1. Iterate over each of the IDs listed by the @idrefs attribute and find the referenced element:
 - 4.1.1. If the referenced element is a partMeta element then check if a @scope attribute is present in the base:
 - 4.1.1.1. If a @scope attribute is not present: a) the <partMeta> content and b) all the <partMeta> metadata properties are governed by the base’s rights expression; they all should be included into a temporary result set. Continue with step 5.
 - 4.1.1.2. If a @scope attribute is present:
 - 4.1.1.2.1. If its value is “riscope:metadata”: only metadata properties are in the scope of this <rightsInfo> element, add only the metadata elements of this <partMeta> element to a temporary result set. Continue with step 5.
 - 4.1.1.2.2. If its value is “riscope:content”: only content is in the scope of this <rightsInfo> element, add only the content described by this <partMeta> element to a temporary result set. Continue with step 5.
 - 4.1.2. If the referenced element is not a <partMeta> element: add the referenced element to a temporary result set. In this case the scope is implied by the element that is referenced and any @scope attribute should be ignored. Continue with step 5.
5. Check the base for an @aspect attribute:
 - 5.1. If an @aspect attribute is not present then all members of the temporary result set should be copied to each of the result sets for the different Rights Aspects.
 - 5.2. If an @aspect attribute is present then all members of the temporary result set should be copied only to the result sets corresponding to the Rights Aspects which are present in the @aspect list.
6. Final step: iterate over the result sets for the different Rights Aspects and interpret the included parts of the content or metadata elements according to the associated aspect. Some members of the result set may not be in a scope specified in the definition of the aspect; such members should be excluded from the result set.

21.5.3 Processing Diagram: Content User view

The following flowchart shows how, from a content user's perspective, the rightInfo associated with a specific element can be obtained.



Key: Refer to previous diagram.
rightsInfo Processing
- v0.0.03, 12/02/10, D J Compton, MediaDev, TR.
- Master in: NFR006

Figure 35: Processing diagram showing how to identify the rights associated with a specific element

21.5.3.1 Content User Flowchart Narrative

1. The goal of the processing: the result will be multiple sets of <rightsInfo> elements which all will apply to this part of the Item. Each of the sets a) correspond to one of the Rights Aspect NewsCodes, and b) may be empty after the processing if no corresponding <rightsInfo> elements were found.
2. Select the part of the Item for which the corresponding <rightsInfo> expression(s) should be determined, this part is the “target” for all subsequent processing steps.



This part must be a) the full content, b) one of the renditions of the content as a whole, c) a part of the content which is described by a <partMeta> element, or d) a single metadata property. The metadata wrappers <itemMeta> or <contentMeta> should not be selected as a target of this processing.

3. Define into which scope of <rightsInfo> elements the target falls:
Match the target against the definitions of corresponding parts for “riscope:content” and “riscope:metadata” of the Rights Info Scope NewsCodes and determine to which scope the target belongs.
Be aware that <partMeta> elements fall under both scopes.
4. Iterate over each <rightsInfo> element which has no @idrefs attribute:
 - 4.1. If a @scope attribute is not present in the <rightsInfo> element then check the <rightsInfo> element against the rules of step 6 and add it to result sets as defined. Earmark the added <rightsInfo> element as “generic scope rightsInfo”. Continue with step 7.
 - 4.2. If a @scope attribute is present and the target falls in the scope of the attribute’s value (see step 3) then check the <rightsInfo> element against the rules of step 6 and add it to result sets as defined. Earmark the added <rightsInfo> element as “generic scope Rights Info”. Continue with step 7.
5. Iterate over each <rightsInfo> element which has an @idrefs attribute that includes the ID of the target:
 - 5.1. If a @scope attribute is not present then check this <rightsInfo> element against the rules of step 6. Be aware that a <rightsInfo> element which is referencing the target by @idrefs overrules <rightsInfo> elements which reference the target by @scope. For that reason if the target should be added to the result set then first delete any <rightsInfo> element which is earmarked as “generic scope Rights Info” from the result set, and then add this <rightsInfo> element. Continue with step 7.
 - 5.2. If a @scope attribute is present and the target falls in the scope of the attribute’s value (see step 3) then check the <rightsInfo> element against the rules of step 6. Be aware that a <rightsInfo> element which is referencing the target by @idrefs overrules <rightsInfo> elements which reference the target by @scope. For that reason if the target should be added to the result set then first delete any <rightsInfo> element which is earmarked as “generic scope Rights Info” from the result set, and then add this <rightsInfo> element. Continue with step 7.
6. Check any @aspect attribute of a <rightsInfo> element:
 - 6.1. If an @aspect attribute is not present then the <rightsInfo> element should be added to the result sets corresponding to each of the Rights Aspect NewsCodes.
 - 6.2. If an @aspect attribute is present then the <rightsInfo> element should be added only to the result sets corresponding to the Rights Aspects which are present in the @aspect list.
7. Final step: iterate over the result sets for the different Rights Aspects and interpret the included parts of the content or metadata elements according to the associated aspect. Some members of the result set may not be in a scope specified in the definition of the aspect; such members should be excluded from the result set.



22 Advanced Metadata Techniques

22.1 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.

22.2 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 @literal or @qcode. 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 @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 (**internationalizationAttributes**):

Name	Datatype
xml:lang	XML Schema language
dir	XML Schema string Enumeration <i>ltr, 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.

22.2.1 Business cases

The Assert wrapper has three main uses:

- ❖ To enable a rich inline representation of concepts used in a NewsML-G2 document, as an alternative to requiring receivers to retrieve remote information about a concept.
- ❖ Where the same concept occurs in multiple properties in a NewsML-G2 document, Assert can be used to merge the details of the concept into a single location, thus making the document less verbose.
- ❖ Ad-hoc concepts that appear only within a single NewsML-G2 Item instance and are not extracted from a concept in a Concept or Knowledge Item

22.2.1.1 Rich inline representation of concepts

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 within a News Item supplementary information, such as how to contact an organisation that requires a richer set of child elements than a property provides.



- ❖ 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:geoArea" qcode="myalias:bn84nh" />
  ...
</contentMeta>
<assert qcode="myalias:bn84nh">
  <geoAreaDetails>
    <position latitude="0.0" longitude="0.0" />
  </geoAreaDetails>
</assert>
```

This example shows how a concept identified by "myalias:bn84nh" which is a Subject of the Item, can be expanded locally using an Assert.

The use of @qcode indicates that the concept is part of a NewsML-G2 controlled vocabulary, and the concept itself is thus globally available to all Items, whereas a @literal may be used as an identifier whose scope is purely local to the Item, or to identify a concept from a vocabulary that is controlled outside the NewsML-G2 framework.

But note that the concept information contained in an Assert in a Item is ALWAYS local in scope, whether the <assert> wrapper is identified by a QCode or a Literal value. 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.

22.2.1.2 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>
  <broader literal="int002" type="cpnatexp:geoAreaCity">
    <name>Amesbury</name>
  </broader>
  <broader literal="int003" type="cpnatexp:geoAreaProvState">
    <name>Wiltshire</name>
  </broader>
  <broader qcode="iso3166-1a3:GBR"26 type="cpnatexp:geoAreaCountry">
```

²⁶ The IPTC recognizes that some well-known CVs, such as those maintained by the International Standards Organization (ISO) are widely used in news exchange. Rather than duplicate this work, the IPTC catalog contains references to these CVs, including the 2-letter and 3 letter country codes. For more information, see the page <http://cvx.iptc.org> hosted by the IPTC.



```

        <name>United Kingdom</name>
    </broader>
</assert>
    
```

22.2.1.2.1 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" /> ← "int014"
  <subject type="cpnat:abstract" qcode="subj:01017001">
    <name>music theatre</name>
  </subject>
  <subject type="cpnat:poi" qcode="artven:int014" /> ← "int014"
</contentMeta>
<assert qcode="artven:int014"> ← "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 → location
      <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 → contact info
      <web>http://www.themet.org</web>
    </contactInfo>
  </POIDetails>
</assert>
    
```

22.2.1.3 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.

22.2.1.3.1 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" /> ← "int014"
  <subject type="cpnat:abstract" qcode="subj:01017001">
    <name>music theatre</name>
  </subject>
  <subject type="cpnat:poi" qcode="artven:int014" /> ← "int014"
</contentMeta>
    
```



```

<assert qcode="artven:int014"> ← "int014"
  <concept>
    <conceptId qcode="artven:int014" /> ← concept ID
    <name>The Metropolitan Opera House</name> ← 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>
        <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="USA">
            <name role="nrol:display">United States</name>
          </country>
          <postalCode>10023</postalCode>
        </address>
      </contactInfo>
    </POIDetails>
  </concept>
</assert>
    
```

22.2.1.3.2 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"> ← @literal
  <concept>
    <conceptId qcode="dummy:091013121256" /> ← "dummy" ID
    <name>The Metropolitan Opera House</name> ← 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/"
  />
  ...
    
```

22.2.2 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>
```

22.3 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.

22.3.1 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.



22.3.2 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 red):

```
<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>
```

22.3.3 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.

22.3.3.1 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.



22.3.3.2 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`).

22.3.3.3 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"/>
...

```



This page intentionally blank



23 Generic Processes and Conventions

23.1 Introduction

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.

23.2 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.

23.3 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”.

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.

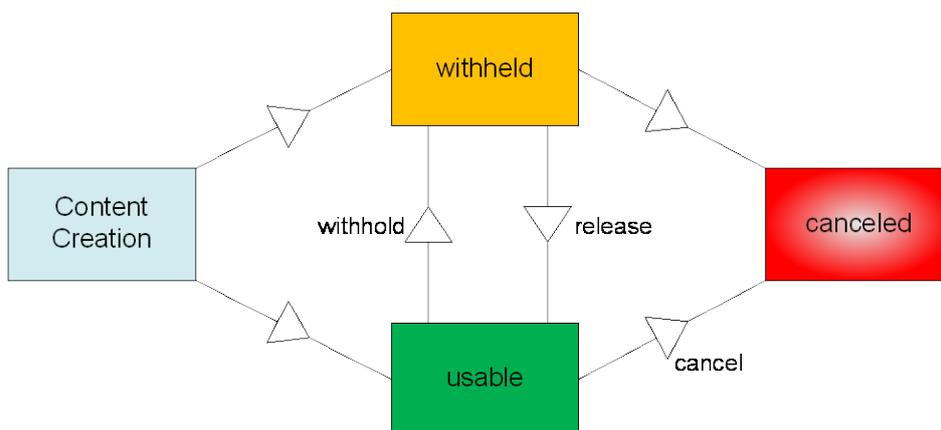


Figure 36: State Transition Diagram for <pubStatus>

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”. For example

```

<newsItem guid="urn:example.com,2009:20090202XYZ999" version="XYZ999" . . . . .>
<catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <itemMeta>
    <itemClass qcode="ni nat: text" />
    <provider qcode="web: www.pressassociation.com" />
    <versionCreated>2009-03-28T11:17:00Z</versionCreated>
    <pubStatus qcode="stat: usable" /> ← publishing status
  </itemMeta>
</newsItem>
  
```



23.4 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 news²⁷, 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.

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.

The change proposes that in these circumstances, the <embargoed> property would be present but empty, and an editorial note would indicate the embargo terms, for example:

```
<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, available at the IPTC Web site www.iptc.org for further information regarding processing rules for <embargoed>. The processing model is illustrated in **Figure 37**.

²⁷ Some public relations organisations impose an “embargo” on news releases to ensure they are used on Sunday, a “slow” news day on which the contents are more likely to be used. Nonetheless, these embargoes are generally respected.

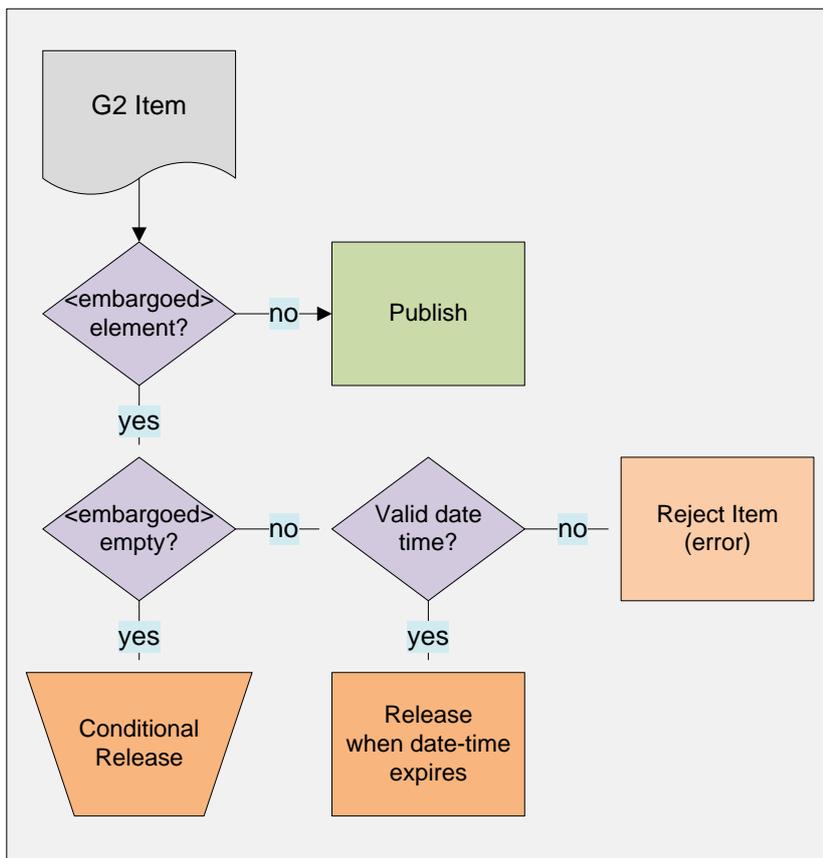


Figure 37: Processing <embargoed>



23.5 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="loc:type: city" qcode="ci ty: 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="ci ty: 28398">  
  <name>Westwood</name>  
  <broader type="cpnat: geoArea" qcode="stat: 3959">  
    <name>Cal i forni a</name>  
  </broader>  
  <broader type="cpnat: geoArea" qcode="country: US">  
    <name>United States</name>  
  </broader>  
</subj ect>
```

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 30 Illustrating Located, Subject and Dateline

```
<contentMeta>  
  <contentCreated>2007- 02- 09T09: 17: 00+03: 00  
  </contentCreated>  
  <located qcode="ci ty: Mosc ow">  
    <broader qcode="cn try: Russi a" />  
  </located>  
  <creator qcode="web: thomsonreuters. com">  
    <name>Thomson Reuters</name>  
  </creator>
```



```
<language tag="en-US" />
<subject type="cpnat:geoArea" qcode="city:Tskhinali">
  <broader qcode="reg:SouthOssetia" >
    <name>South Ossetia</name>
  </broader>
</subject>
<subject type="cpnat:geoArea" qcode="city:Tbilisi">
  <broader qcode="cntry:Georgia" >
    <name>Georgia</name>
  </broader>
</subject>
<dateline>MOSCOW, Monday (Reuters)</dateline>
</contentMeta>
```

[<<return to Code Listings index](#)



23.6 Processing Updates and Corrections

By its nature, news may need frequent updating, and in some cases correcting, as new facts come to light. 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 an <edNote> which can be displayed to inform recipients of the reason for the update.

NewsML-G2 provides 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 that the rules for applying the codes will be up to the provider to specify as part of their NewsML-G2 implementation so that their end-users can correctly process the instruction.

The Scheme URI for the IPTC Signal NewsCode is <http://cv.iptc.org/newscodes/signal> and the recommended Schema Alias is “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.	This 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.	This 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 guid="tag:afp.com,2008:TX-PAR:20090529:JYC80" ←———— current item
  version="2" ←———— new version
  ....>
  <itemMeta>
    ....
    <edNote>Update previous version by appending these paragraphs</edNote>
    <signal qcode="sig:update" />
    ....
  </itemMeta>
  ....
</newsItem>
    
```

The impact of the change can also be signalled by the optional @severity property of <signal>, which uses a recommended IPTC Severity NewsCode. The Scheme URI is <http://cv.iptc.org/newscodes/severity> and a recommended Scheme Alias of “sev”. There are two members of the CV:

Code	Definition
minor	A minor impact
major	A major impact.

23.6.1 Example Use Case using <link> to previous version and @severity

Some content systems do not provide a common identifier for updated content and previous versions of the same content. In these cases, it may not be possible to use GUID to identify previous versions, and the provider would need to use <link> to refer to the GUID of a previous version, using @residref.



The relationship between the current item and the item referenced by <link> is expressed using @rel. The appropriate code value from the IPTC Item Relation NewsCode would be “previousVersion”:

```
<newsItem guid="tag:afp.com,2008:TX-PAR:20090529:JYC85" ← new item
  version="1" ← first version
  ....
  <itemMeta>
    ....
    <edNote>
      Replaces previous version. MUST correction, updates name of minister
    </edNote>
    <signal severity="sev:major" qcode="sig:correction" /> ← severity
    <link
      rel="irel:previousVersion"
      residref="tag:afp.com,2008:TX-PAR:20090529:JYC80" ← previous item
      version="1" /> ← first version
    ....
  </itemMeta>
  ....
</newsItem>
```



23.7 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:cwarn" />
```

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 Scheme Alias is “cwarn”. The Scheme values are:

Code	Definition
death	The content could be perceived as offensive due to the discussion or 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:cwarn" />  
<exclAudience qcode="cwarn:nudity" />  
<exclAudience qcode="cwarn:language" />
```



This page intentionally blank



24 Verification and Security

24.1 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 to maintain document/content security.

24.2 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 conveying 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 property 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 property 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



Code	Definition
	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 the @target property) 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 31 Hop History

```
<?xml version="1.0" encoding="UTF-8" ?>
<newsItem
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    XSD/NewsML-G2_2.9-spec-All-Core.xsd"
  guid="urn:newsmml:acmenews.com:20111125T1205:HOP-HISTORY-EXAMPLE"
  version="3"
  standard="NewsML-G2"
  standardversion="2.9" xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_16.xml" />
  <catalogRef href="http://www.acmenews.com/synd/catalogs/anncodes.xml" />
  <hopHistory>
    <hop seq="1" timestamp="2010-11-10T16:15:00Z">
      <party qcode="comp:TR" />
      <!-- alternative representation -->
      <party literal="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:acquiremedia" />
      <!-- 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>
</newsItem>
```



```
<rightsInfo>
  <copyrightHolder literal="Acme News and Media LLC" />
</rightsInfo>
<itemMeta>
  <itemClass qcode="inat:text" />
  <provider literal="ANM" />
  <versionCreated>2011-11-12T20:15:00Z
</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<!-- metadata and content -->
<!-- metadata and content -->
<!-- metadata and content -->
</newsItem>
```

[<<return to Code Listings index](#)

24.3 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:

- ❖ Hash Type NewsCodes; Scheme URI <http://cv.iptc.org/newscodes/hashtype/> with a recommended scheme alias of “htype”. Scheme values are:

Code	Definition
MD5	The MD5 hash function
SHA-1	The SHA-1 hash function
SHA-2	The SHA-2 hash function

- ❖ 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.



25 Changes to G2-Standards Family

25.1 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.iptc.org). The latest changes are in the topmost section, followed by earlier changes in descending order.

25.2 NewsML-G2 v2.7-EventsML-G2 1.6 →NewsML-G2/EventsML-G2 2.9

25.2.1 Unification of NewsML-G2 and EventsML-G2

Please refer to [Section 3](#) on [page 19](#).

25.2.2 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.

25.2.3 Hop History

Add a machine-readable transaction log an Item. See [24.2](#)

25.2.4 Concept Reference

Enable Event Concepts to be referenced within NewsML-G2 Packages. See [10.3.3](#) and [14.5.3.1](#)

25.2.5 Extend News Message Header

Enable News Messages to express more semantically-rich properties using QCode values. See [17.2.1](#)

25.2.6 Hash value

Enable the receiver of a NewsML-G2 Item to determine whether the content has been altered. See [0](#)

25.2.7 Extend Icon attributes

Add properties of <icon> to enable a basis for choice when multiple icons for content are available. See [8.2.2.12](#)

25.2.8 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 [9.2.2](#)

25.2.9 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 [9.2.2](#)



25.2.10 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 [9.2.2](#)

25.2.11 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 [9.2.2](#)

25.2.12 Content Warning Best Practice

Add Best Practice for expressing a warning about content and for adding refinements to the warning.

25.2.13 Expressing updates and corrections

Best Practice and structures for expressing that a previous Item has been updated or corrected by the received version

25.3 News Architecture (NAR) v1.7 → v1.8

25.3.1 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](#).

25.3.2 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.

25.3.3 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.

25.3.4 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.

25.3.5 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 [Expressing quantitative values using <related>](#)

25.3.6 Extend <contentMeta> for Concept Items and Knowledge Items

This change creates a group of core descriptive metadata properties for G2 items and use it 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

25.3.7 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>.



25.4 NewsML-G2 v2.6 → v2.7

No specific changes; inherits all appropriate changes from NAR v1.8

25.5 EventsML-G2 v1.5 → v1.6

No specific changes; inherits all appropriate changes from NAR v1.8

25.6 News Architecture (NAR) v1.6 → v1.7

25.6.1 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 <partMeta>](#)

25.6.2 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](#).

25.7 NewsML-G2 v2.5 → v2.6

No specific changes; inherits all appropriate changes from NAR v1.7

25.8 EventsML-G2 v1.4 → v1.5

No specific changes; inherits all appropriate changes from NAR v1.7

25.9 News Architecture (NAR) v1.5 → v1.6

25.9.1 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>
```

25.9.2 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.

25.9.3 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.

25.9.4 Add `<sameAs>` to the scheme declarations of a catalog

This feature is fully documented in [13.7 Private versions and extensions of CVs](#)

25.10 NewsML-G2 v2.4 → v2.5

No specific changes; inherits all appropriate changes from NAR v1.6

25.11 EventsML-G2 v1.3 → v1.4

No specific changes; inherits all appropriate changes from NAR v1.6

25.12 News Architecture (NAR) v1.4 → v1.5

The following changes are inherited by NewsML-G2 2.4 and EventsML-G2 1.3

25.12.1 @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.

25.12.2 <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.

25.12.3 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.

25.12.4 Scheme Code Encoding

A full processing model for Scheme URIs and QCodes was defined. See [13.8](#)

25.12.5 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>

25.12.6 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.

25.12.7 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>
```



25.12.8 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.

25.12.9 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.

25.13 NewsML-G2 v2.3 → v2.4

Inherits changes from NAR v1.5 plus the following changes specific to NewsML-G2 2.4

25.13.1 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"
/>
```

25.14 EventsML-G2 v1.2 → v1.3

Inherits the changes from NAR v1.5. No other changes.

25.15 News Architecture (NAR) 1.3 → 1.4

The following changes are inherited by NewsML-G2 2.3 and EventsML-G2 1.2



25.15.1 Revised Embargo

An embargo can now have an undefined date and time. See [Embargo](#)

25.15.2 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 [Supplementary information about a Concept](#)

25.16 NewsML-G2 v2.2 → v2.3

Inherits the changes described in [25.2](#). plus the following changes specific to NewsML-G2 2.3

25.16.1 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")

25.16.2 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 [9.2.1.10](#)

25.16.3 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 [9.2.2.1](#)

25.17 EventsML-G2 v1.1 → v1.2

Inherits the changes from NAR v1.4. No other changes

25.18 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>.



26 Additional Resources

The IPTC web site www.iptc.org has a wealth of resources for implementers. The site is divided into logical areas of interest using horizontal menus under the site banner:



Following the link to News Exchange Formats displays menu choices for each IPTC Standard:



And under each of the Standards listed is a consistent menu of available resources:



The "Specification" page for each Standard contains a link to download a zip archive of resources, including, for the G2 Standards, all necessary XML Schema files, documentation and examples.