



## **IPTC Standards**



## **Guide for Implementers**

Public Release

Document Revision 2

International Press Telecommunications Council  
Copyright © 2009. All Rights Reserved  
[www.iptc.org](http://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 © 2009 IPTC, the International Press Telecommunications Council. All Rights Reserved.

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.

This document is published under a **License Agreement**. If you do not agree to the terms 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](http://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](http://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**.



## 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
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
SERGEANT, 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:

- ❖ EventsML-G2, version 1.3
- ❖ NewsML-G2, version 2.4
- ❖ SportsML-G2, version 2.1
- ❖ The News Architecture upon which these Standards are based is version 1.5

## 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

## 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="2.2"
```

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="2.2"
```



Indicates especially  
important notes or  
cautions to  
implementers

## Note on Spelling (English)

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

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



## Table of Contents

Introduction .....	2
Purpose and Audience .....	2
Terms of Use.....	2
Acknowledgments .....	3
About the G2 Standards.....	3
Document History .....	3
Conventions used by this document.....	4
Note on Spelling (English).....	4
<b>Table of Figures</b> .....	12
<b>Table of Code Listings</b> .....	13
1 Executive Summary.....	15
1.1 Why standards matter .....	15
1.2 What Are the G2 Standards? .....	15
1.3 Business Drivers.....	15
1.4 Business Requirements .....	15
1.5 G2 Design and Benefits .....	16
1.6 The role of the IPTC.....	16
2 Advice on using the Guide .....	17
2.1 Note on Code Listings .....	17
3 How News Happens .....	19
3.1 Introduction .....	19
3.2 Assignment .....	19
3.3 Information Gathering.....	19
3.4 Verification .....	20
3.5 Dissemination.....	20
3.6 Archiving.....	20
4 Anatomy of G2.....	23
4.1 Data model.....	23
4.2 Metadata.....	24
4.2.1 G2 Properties.....	24
4.2.2 Controlled Values .....	25
4.2.3 Concepts.....	25
4.2.4 IPTC NewsCodes .....	25
4.2.5 Qualified Code schemes (QCodes) .....	26
4.3 Conformance Levels.....	27
4.4 Introductory example: a simple News Item.....	28
4.4.1 The top level <newsItem> .....	29
4.4.2 Item Metadata <itemMeta> .....	31



4.4.3	Content Metadata <contentMeta> .....	31
4.4.4	The Content <contentSet> .....	32
5	Text .....	33
5.1	Introduction .....	33
5.2	Example .....	33
5.3	The top level .....	33
5.3.1	<newsItem> .....	33
5.3.2	Catalogs .....	34
5.3.3	Rights information.....	35
5.3.4	Summary.....	35
5.4	Item Metadata.....	35
5.4.1	<itemClass> .....	36
5.4.2	<provider> .....	36
5.4.3	<versionCreated>.....	36
5.4.4	Optional Management Metadata elements .....	36
5.4.5	Links.....	37
5.4.6	Summary.....	37
5.5	Content Metadata .....	37
5.6	The Content.....	38
5.7	Putting it together.....	39
5.8	Other content options .....	40
5.8.1	Inline XML.....	40
5.8.2	Inline data .....	40
6	Pictures and Graphics .....	43
6.1	Introduction.....	43
6.1.1	Use Case.....	43
6.2	Photo metadata.....	43
6.2.2	Example Metadata .....	44
6.3	Photo data.....	47
6.3.1	Remote Content wrapper .....	47
6.3.2	News Content Attributes .....	48
6.3.3	Target Resource Attributes .....	48
6.3.4	News Content Characteristics .....	49
6.4	The Complete Picture .....	50
7	Audio and Video .....	53
7.1	Introduction .....	53
7.2	Use Case 1: Simple Video for Broadcast .....	53
7.2.1	Metadata.....	53
7.2.2	Content .....	57
7.3	Use case 2: Multiple renditions of web video and audio .....	60
8	Packages.....	61



8.1	Introduction.....	61
8.2	Conveying Package Structures .....	63
8.3	Simple Package Structure .....	63
8.3.1	Group Set <groupSet>.....	64
8.3.2	Item Reference <itemRef> .....	64
8.4	Hierarchical Package Structure.....	66
8.5	List Type Package Structure (ordered, bag, alternative).....	68
8.6	Package Processing Considerations .....	70
8.6.1	Other G2 Items.....	70
8.6.2	Facilitating the Exchange of Packages.....	71
8.6.3	A “Top Ten” Package Example.....	72
8.6.4	Packaging non-G2 Resources.....	74
8.7	Packages and Links.....	74
8.7.1	The difference in a nutshell .....	74
8.7.2	Purpose and features of the <packageItem> .....	74
8.7.3	When to use <link>.....	75
8.7.4	Link properties.....	75
8.7.5	Link examples .....	77
9	Concepts and Concept Items.....	79
9.1	Introduction.....	79
9.2	What is a Concept? .....	79
9.3	Concept Item.....	79
9.4	Creating Concepts – the <concept> element.....	80
9.4.1	Concept ID <conceptId> .....	80
9.4.2	Concept Name <name> .....	80
9.4.3	Concept Type <type> and Facet <facet> .....	81
9.4.4	Concept Definition <definition> .....	81
9.4.5	Note <note>.....	82
9.4.6	Remote Info <remoteInfo> .....	82
9.5	Relationships between Concepts .....	82
9.6	Entity Details.....	83
9.6.1	Person Details <personDetails> .....	83
9.6.2	Organisation Details <organisationDetails> .....	85
9.6.3	Geopolitical Area Details .....	85
9.6.4	Point of Interest <poiDetails> .....	85
9.6.5	Object Details <objectDetails> .....	86
9.7	Code listings .....	87
9.8	Supplementary information about a Concept.....	89
9.9	Concepts and Controlled Values .....	90
9.9.1	Concept Identifier and Concept URI .....	90
9.9.2	From Concept URI to QCodes .....	91



9.9.3	Resolution of QCodes.....	91
9.10	Concepts in Practice .....	92
10	Knowledge Items .....	93
10.1	Introduction .....	93
10.2	Structure and Properties.....	94
10.2.1	The <knowledgeItem> element.....	95
10.2.2	Management Metadata.....	95
10.2.3	Content Metadata.....	96
10.2.4	Concept Set .....	96
10.3	Use Case: a Controlled Vocabulary for <accessStatus> .....	97
11	Controlled Vocabularies and QCodes .....	101
11.1	Introduction .....	101
11.2	Business Case.....	101
11.3	How QCodes work .....	101
11.3.1	Scheme Alias to Scheme URI .....	102
11.3.2	Concept URI.....	102
11.4	QCodes and Taxonomies .....	104
11.5	Managing Controlled Vocabularies as G2 Schemes.....	105
11.5.1	Knowledge Items .....	105
11.5.2	Creating a new Scheme .....	105
11.5.3	Creating a new Knowledge Item for distributing a CV .....	105
11.5.4	Maintaining a CV .....	106
11.6	Processing Controlled Vocabularies.....	106
11.6.1	Resolving Scheme Aliases.....	107
11.6.2	Resolving Concept URIs .....	107
11.7	Private versions of CVs .....	108
11.7.1	Example Business Case .....	108
11.7.2	Solution.....	108
11.8	Complementary CVs.....	109
11.9	QCodes and non-URI characters.....	110
11.9.1	Non-ASCII characters.....	110
11.9.2	White Space in Codes.....	111
11.10	Syntactic Processing of QCodes .....	111
11.10.1	Creating QCodes from Scheme Aliases and Codes .....	111
11.10.2	Processing Received Codes.....	111
11.11	QCode and Literal Identifiers.....	112
11.11.1	The difference in a nutshell.....	112
11.11.2	What @literal is; what it is NOT .....	112
11.11.3	Use of @literal in a G2 Item.....	112
12	EventsML-G2 .....	115
12.1	A standard for exchanging news event information .....	115





12.1.1	Business Advantages of adopting EventsML-G2.....	115
12.1.2	Structure and compatibility .....	116
12.2	Event Information – What, Where, When and Who .....	118
12.2.1	What is the Event?.....	118
12.2.2	Where does it take place? .....	118
12.2.3	When does it take place? .....	118
12.2.4	Who is involved? .....	118
12.3	Event Coverage .....	118
12.4	Event Properties .....	119
12.4.1	Event Description .....	119
12.4.2	Event Relationships.....	120
12.4.3	Event Details Group.....	122
12.5	Event Use Cases .....	134
12.5.1	Events in a News Item .....	134
12.5.2	Event as a Concept.....	136
12.5.3	Multiple Event Concepts in a Knowledge Item.....	140
13	SportsML-G2 .....	143
13.1	Introduction .....	143
13.2	Business Advantages .....	143
13.3	Structure .....	143
13.3.1	XML Schema and Catalogs .....	144
13.3.2	Item Metadata.....	145
13.3.3	Content Metadata.....	145
13.3.4	InlineXML and the Sports Content component.....	146
13.3.5	Sports Event wrapper in more detail.....	147
13.3.6	SportsML-G2 Packages.....	152
14	Exchanging News: News Messages.....	159
14.1	Introduction .....	159
14.2	Structure .....	159
14.2.1	News Message Header <header> .....	160
14.2.2	News Message payload <itemSet> .....	161
15	Migrating IPTC 7901 to G2.....	163
15.1	Introduction .....	163
15.1.1	Message Header .....	163
15.1.2	Message Text .....	164
15.1.3	Post-Text Information.....	164
16	News Industry Text Format (NITF).....	167
16.1	Introduction .....	167
16.2	Overview of NITF and G2 equivalents.....	168
16.2.1	Root element <nitf> .....	168
16.2.2	Document metadata <head> .....	168



16.3	Content metadata <body.head> .....	170
16.4	Body Content <body.content> .....	171
16.5	End of Body <body.end> .....	171
17	Mapping Embedded Photo Metadata to G2.....	173
17.1	Introduction .....	173
17.2	IPTC Metadata for XMP .....	173
17.3	Synchronizing XMP and legacy metadata.....	174
17.4	Picture services using IIM/XMP .....	174
17.5	Rationale for moving to G2.....	174
17.6	IIM resources .....	175
17.7	Approach .....	175
17.8	IIM to G2 Field Mapping Reference .....	175
17.9	IIM to G2 Mapping Example.....	176
17.9.1	Object Name / Title.....	177
17.9.2	Urgency.....	177
17.9.3	Category, Supplementary Category.....	178
17.9.4	Keywords.....	178
17.9.5	Special Instruction.....	178
17.9.6	Date Created, Time Created.....	179
17.9.7	By-line, Credit, Source.....	179
17.9.8	City, Province/State and Country .....	180
17.9.9	Original Transmission Reference.....	181
17.9.10	Headline .....	181
17.9.11	Copyright Notice .....	182
17.9.12	Caption/Abstract.....	182
17.9.13	Writer/Editor.....	182
17.10	Reconciling G2 with embedded metadata.....	183
18	Advanced Metadata Techniques .....	185
18.1	Introduction .....	185
18.2	The Assert wrapper .....	185
18.2.1	Business cases .....	185
18.2.2	Using Assert to map from IIM or XMP.....	189
18.3	Inline Reference.....	189
18.3.1	Quantify Attributes Group .....	189
18.3.2	Linking text content to an Inline Reference.....	189
18.3.3	Using <inlineRef> and <assert> together .....	190
19	Generic Processes and Conventions.....	193
19.1	Introduction .....	193
19.2	Processing Rules for G2 Items .....	193
19.3	Publishing Status .....	193
19.4	Embargo.....	195



19.5	Geographical Location .....	197
19.6	Processing Updates and Corrections.....	199
20	Changes to G2-Standards .....	201
20.1	Introduction .....	201
20.2	News Architecture (NAR) 1.3 → 1.4 .....	201
20.2.1	Revised Embargo .....	201
20.2.2	New Remote Info element .....	201
20.3	NewsML-G2 v2.2 → v2.3.....	201
20.3.1	Add <role> to partMeta. ....	201
20.3.2	Revised Time Delimiter.....	201
20.3.3	Revised @duration property and new @durationunit. ....	201
20.4	EventsML-G2 v1.1 → v1.2.....	201
20.5	SportsML-G2 v2.0 → v2.1.....	201
20.6	News Architecture (NAR) v1.4 → v1.5 .....	202
20.6.1	@rendition .....	202
20.6.2	<assert>.....	202
20.6.3	Hint and Extension Point.....	202
20.6.4	Scheme Code Encoding .....	202
20.6.5	Add @rendition to the icon property.....	202
20.6.6	Ranking Multiple Elements .....	202
20.6.7	Keyword property .....	203
20.6.8	Multiple Generators .....	203
20.6.9	Cardinality of Icon.....	203
20.7	NewsML-G2 v2.3 → v2.4.....	203
20.7.1	New dimension unit indicators for visual content .....	203
20.8	EventsML-G2 v1.2 → v1.3.....	204
21	Additional Resources.....	205
	License Agreement .....	207



## Table of Figures

Figure 1: The heart of 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 .....	16
Figure 2: How the NAR builds into NewsML-G2, EventsML-G2 and SportsML-G2.....	23
Figure 3: Browser view of the remote Catalog.....	26
Figure 4: NewsML-G2 News Item derived from the News Architecture .....	28
Figure 5: One of the IPTC Custom metadata panels available in Adobe Photoshop CS.....	43
Figure 6: A "Top Ten" package on the Web    Source: Thomson Reuters.....	61
Figure 7; Structural model of a NewsML-G2 Package Item resembles the News Item.....	62
Figure 8: Simple Package using Item References <itemRef> .....	64
Figure 9: Simple Package Relationship.....	65
Figure 10: Code outline of Hierarchical Package with two Groups.....	67
Figure 11: Relationship diagram of Hierarchical Package.....	67
Figure 12: Using <groupRef> to create an ordered package of resources .....	69
Figure 13: Groups G1 and G2 are both children of the Root Group.....	69
Figure 14: Diagrams of Package Profiles. The numbers in brackets indicate the required items .....	71
Figure 15: Concept Items share the same template as News Items and Package Items.....	80
Figure 16: Entering the Concept URI in a browser returns human-readable information.....	91
Figure 17: Information Flow for Concepts and Knowledge Items .....	93
Figure 18: Structure of a Knowledge Item mirrors that of other G2 Items .....	95
Figure 19: Human-readable browser page of a Concept URI.....	103
Figure 20: Multiple event information carried as <events> in a News Item.....	117
Figure 21: An event conveyed as a Concept using <eventDetails> .....	117
Figure 22: Hierarchy of Events created using Event Relationships.....	120
Figure 23: An event represented in a News Item or Concept Item uses a common structure.....	136
Figure 24: Structure of a SportsML-G2 document matches NewsML-G2 .....	144
Figure 25: Diagram of Package Group structure.....	155
Figure 26: News Messages can convey any number of any kind of G2 Item .....	159
Figure 27: Syndication workflow using Destination and Channel .....	161
Figure 28: One of the IPTC Core metadata panels from an image opened in Adobe Photoshop CS2. ...	173
Figure 29: Opening File Info panel of the image ( <i>photo and metadata ©Thomson Reuters</i> ).....	176
Figure 30: State Transition Diagram for <pubStatus> .....	194
Figure 31: Processing model for <embargoed> .....	195
Figure 32: Processing <embargoed> from NewsML-G2 2.3/EventsML-G2 1.2.....	196



## Table of Code Listings

LISTING 1: Example of a NEWSML-G2 Document .....	29
LISTING 2: A NEWSML-G2 Document .....	39
LISTING 3: Conveying a Picture in NewsML-G2 .....	50
LISTING 4: Simple Video in NewsML-G2 .....	57
LISTING 5: Multiple renditions of audio/video .....	60
LISTING 6: Simple Group Set at CCL .....	65
LISTING 7: Simple Group Set at PCL .....	66
LISTING 8: Hierarchical Package .....	67
LISTING 9: List Type Package .....	70
LISTING 10: The complete Top Ten style News Package .....	72
LISTING 11: Simple Group Set using @href at CCL .....	74
LISTING 12: "Abstract" Concept .....	87
LISTING 13: "Named Entity" Concept .....	88
LISTING 14: Knowledge Item for Access Codes .....	98
LISTING 15: A Set of Events carried in a News Item .....	135
LISTING 16: Event sent as a Concept Item .....	138
LISTING 17: Update to the Previously-Sent Event .....	139
LISTING 18: Two Related Events conveyed in a Knowledge Item .....	141
LISTING 19: A complete sample SportsML-G2 document .....	150
LISTING 20: Sports story in NewsML-G2/NITF .....	152
LISTING 21: SportsML-G2 Package .....	156
LISTING 22: News Message conveying a complete News Package .....	162
LISTING 23: An NITF marked-up article conveyed in G2 <inlineXML> .....	171
LISTING 24: Embedded photo metadata fields mapped to NewsML-G2 .....	182
LISTING 25: Illustrating Located, Subject and Dateline .....	197



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 What Are the G2 Standards?

The G2 family allows the exchange of ANY kind of news content, be it text, pictures, audio or video, and in ANY format. Information partners are free to choose any format that they need because the G2 family is content-neutral and does not specify the mark-up of text or encoding of binary content.

Text mark-up such as NITF (News Industry Text Format, an IPTC XML standard), and XHTML are being used with G2. Other standards such as Atom and RSS, and microformats such as hNews, are all compatible with NewsML-G2. EventsML-G2 is compatible with the iCal standard.

G2 models the way that professional news organisation work, but the standards go 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. Using G2, news organisations can become part of the Semantic Web, opening up new business opportunities in the digital marketplace.

■ *"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**

## 1.3 Business Drivers

Most business issues faced by media organisations are related to the growth of the Internet, which has not only increased the availability of news content, but created new, and arguably, better ways of consuming 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 enterprise systems.

## 1.4 Business Requirements

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

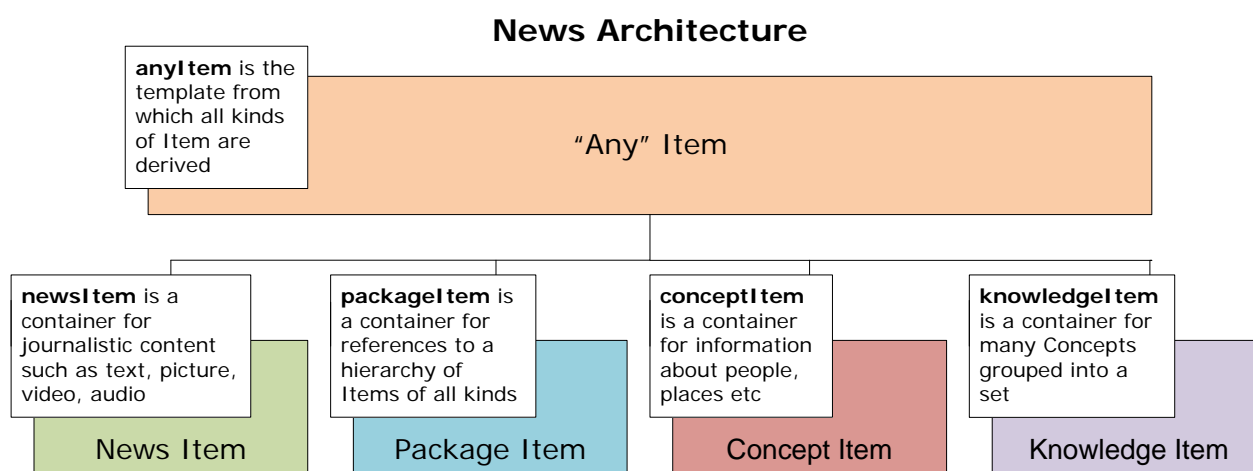
- ❖ Fit an MMM strategy (Multi-media, Multi-channel, Multi-platform)
- ❖ Handle texts, pictures, graphics, animated, audio or video news
- ❖ Be a lightweight container for news, simple to implement and extend, yet offer powerful features for advanced applications.
- ❖ Be useful at all stages of the lifecycle of news, from initial event planning, through content gathering, syndication, to archiving.

## 1.5 G2 Design and Benefits

All 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 the G2-Standards 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 G2 standards enable 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 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 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**

There are currently three G2 standards using the NAR framework:

- ❖ **NewsML-G2** – a wrapper for multi-media news content, from single news items to structured packages containing multiple kinds of related content (e.g. text, pictures, video, Flash)
- ❖ **EventsML-G2** – a standard for describing news coverage, including calendar items and breaking news.
- ❖ **SportsML-G2** – sports results and statistics using a plug-in architecture for recording highly-detailed actions for specific sports (e.g. ice hockey, football, baseball)

## 1.6 The role of 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.

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

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.



## 2 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 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** 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 G2**. It may also be helpful to read the Chapter on **Text** since this introduces some of the core structures used throughout G2.

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 **Advanced Metadata Techniques** and **Generic Processes and Conventions**

There are chapters dedicated to using 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 G2 is the ability to convey “packages” of all types of media in managed relationships. This is detailed in a Chapter on **Packages**.

Implementers who are migrating from IPTC 7901 or NITF will find specific information in **Migrating IPTC 7901 to G2, News Industry Text Format (NITF). Mapping Embedded Photo Metadata to G2** contains advice for mapping IIM and embedded application metadata, such as Photoshop XMP, to 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**.

A brief history of **Changes to G2-Standards** summarizes new features added since the last Revision of the Guidelines

### 2.1 Note on Code Listings

All of the Code Listings have been validated against the appropriate XML Schema(s). 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](http://www.iptc.org). Follow the links to News Exchange Formats → NewsML-G2 (or EventsML-G2, or SportsML-G2) → Specification to obtain a link to download a ZIP package of schema files, documentation and examples.

Individual Schema Files used in this document may be downloaded from:

- ❖ <http://www.iptc.org/std/EventsML-G2/1.3/specification/>
- ❖ <http://www.iptc.org/std/NewsML-G2/2.4/specification/>

For SportsML-G2, a ZIP package of schema, documentation and examples is at <http://www.iptc.org/std/SportsML-G2/2.1.zip>



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.



This page intentionally blank

## 3 How News Happens

### 3.1 Introduction

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

Events become 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

### 3.2 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** is designed to address.

### 3.3 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.



### 3.4 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.

### 3.5 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, most organisations currently use a human-readable message to indicate rights.

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.

### 3.6 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 is almost literally meaningless.



This page intentionally blank

## 4 Anatomy of G2

### 4.1 Data model

The three related G2 standards, NewsML-G2, EventsML-G2, and SportsML-G2, are specific implementations in XML of the News Architecture. As was shown in [Figure 1](#) the abstract class anyItem sits at the top of the NAR hierarchy, with four derived classes, newsItem, packageItem, conceptItem and knowledgeItem, sitting beneath it.

These components are used to build NewsML-G2 or EventsML-G2 documents, as shown below:

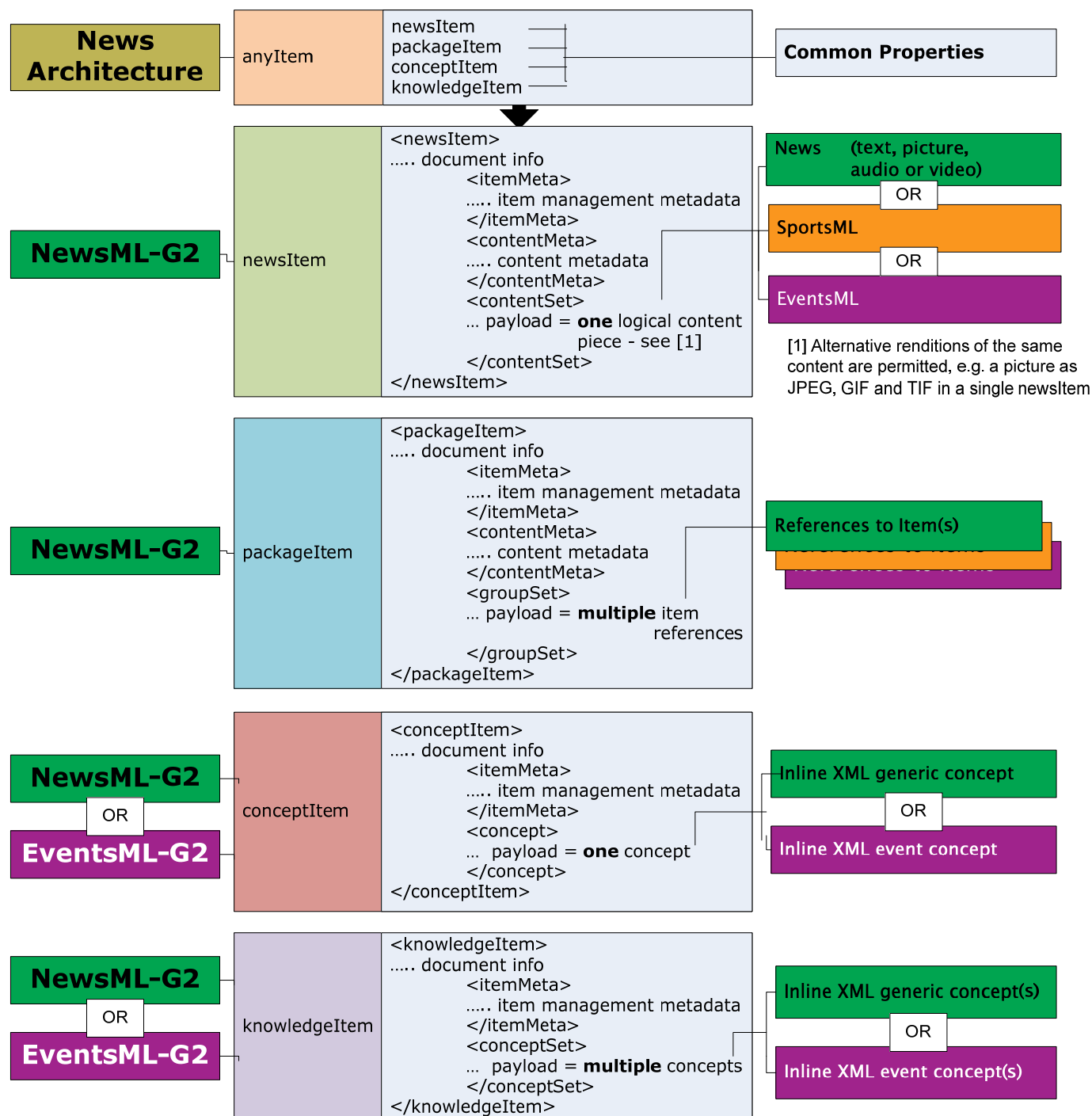


Figure 2: How the NAR builds into NewsML-G2, EventsML-G2 and SportsML-G2



SportsML-G2 is always conveyed in a NewsML-G2 document structure, whereas EventsML-G2 can also be “standalone” by using the conceptItem to express News Event information. We look in more detail at this implementation of **EventsML-G2**.

The News Architecture itself is independent of any specific technical implementation. This Guide describes the implementation of the IPTC-provided XML Schema, but the model could be implemented using object-oriented software, for example Java.

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

## 4.2 Metadata

One of the rules of 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 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*

In a digital environment, this rudimentary metadata is no longer sufficient: the volume of content available means we must have more metadata in order to 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. G2 Properties offer powerful support for metadata that can be used to create links to the content.

### 4.2.1 G2 Properties

A core design feature of 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, 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 G2 property is based on one of a number of re-usable templates

#### 4.2.1.1 String, Block and Label Type

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

#### 4.2.1.2 Qualified Properties

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

#### 4.2.1.3 Flexible Properties

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



#### 4.2.1.4 Other Property Types

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

#### 4.2.1.5 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.

### 4.2.2 **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. G2 provides this functionality in a lightweight and portable manner.

### 4.2.3 **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 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](#).

### 4.2.4 **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 providers 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”<sup>1</sup> it may not be able to process the content correctly. G2 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 G2 message would be:

```
<itemclass qcode="nature:picture" />
```

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

---

<sup>1</sup> Greek for “photograph”

## 4.2.5 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= **newsItem Nature**). We refer to these aliases throughout G2 documentation as **scheme aliases**. The second part of the QCode “picture” is a reference into the newsItem nature vocabulary.

Scheme aliases are resolved by looking in **Catalog** information carried at the root level of a 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:

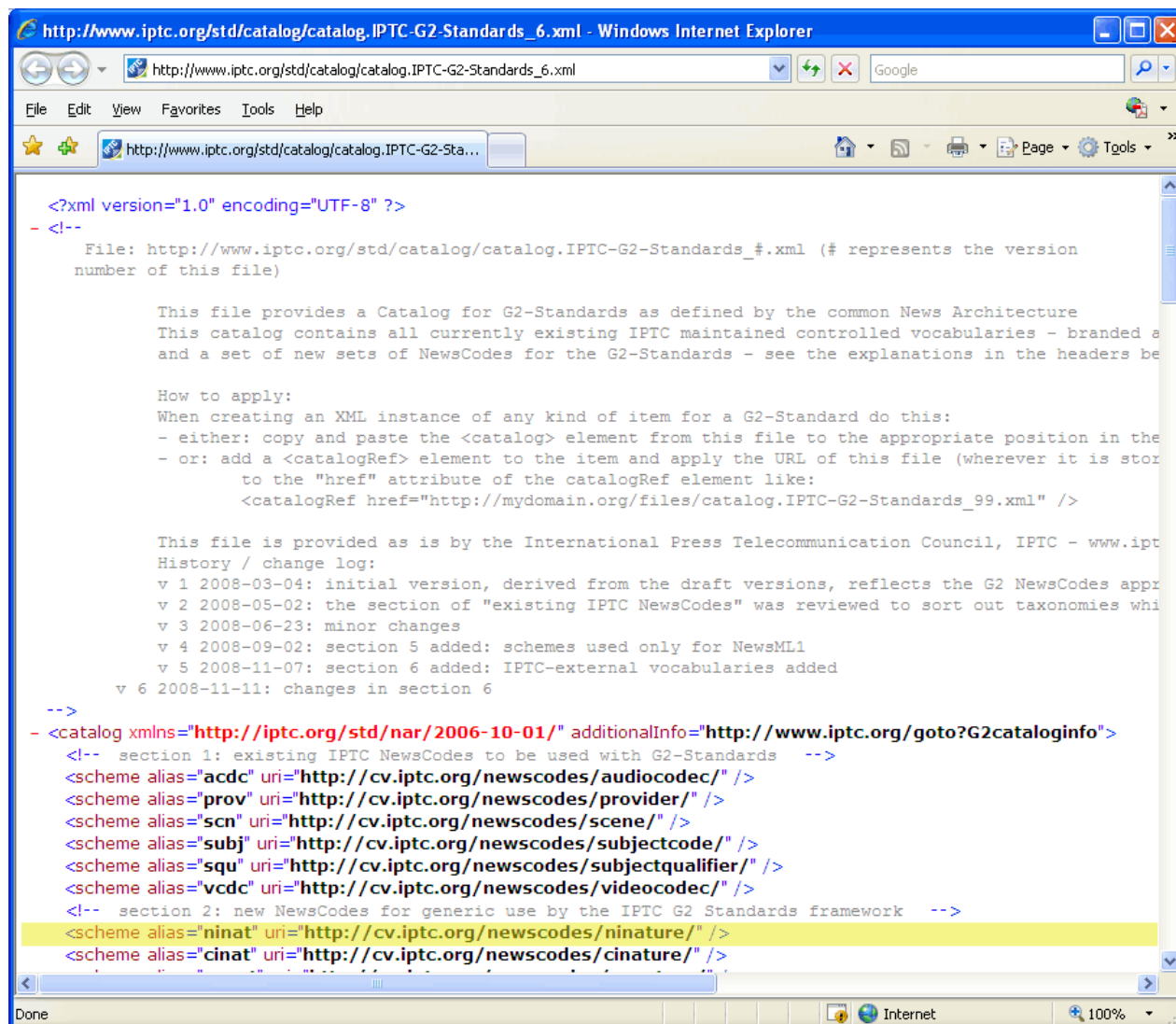


Figure 3: Browser view of the remote Catalog

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 human-readable form that the allowed terms (in English) in the controlled vocabulary<sup>2</sup> are as follows:

<sup>2</sup> Note that this particular piece of information does not convey the format or encoding of the content, which is described separately.

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

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 G2 standards require.

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

#### 4.2.5.1 QCodes and QCode types

QCodes are used throughout G2 documents as a way of expressing controlled values. Apart from the @qcode attribute used by many properties, other attributes of G2 properties 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".

#### 4.2.5.2 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">
```

### 4.3 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.

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

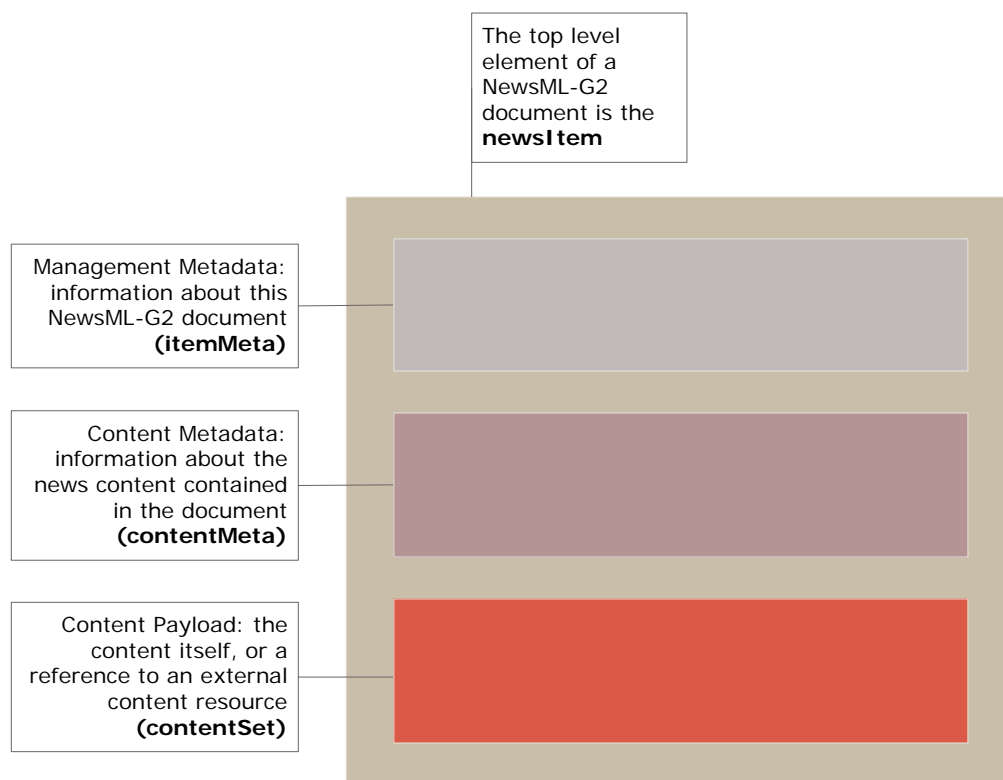
If a G2 document implements PCL features, it MUST declare the conformance level in the root level of the Item. If the conformance property is missing, CCL is assumed.:

```
conformance="power"
```

## 4.4 Introductory example: a simple News Item

The G2 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 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.** The **<itemMeta>** structure conveys information 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.** The **<contentMeta>** structure contains information 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](http://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:20091107: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/
    NewsML-G2_2.4-spec-NewsItem-Core.xsd"
  standard="NewsML-G2"
  standardversion="2.4"
  xml:lang="en-GB" >
  <catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_11.xml"
/>

  <itemMeta>
    <itemClass qcode="ni:at: text" />
    <provider literal="IPTC" />
    <versionCreated>2009-11-07T12:12:12+01:00</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>

  <contentMeta>
    <contentCreated>2009-11-06</contentCreated>
    <located literal="Washington" />
    <creator literal="LLeMeur">
      <name>Laurent Le Meur</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>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 2.4 approved by the IPTC in Washington</h1>
          <p>
            <span class="dateline">Washington, November 6 (IPTC) -</span>
            The International Press Telecommunications Council Standards Committee
            today approved the latest version of its News Exchange Format
            NewsML-G2.
          </p>
        </body>
      </html>
    </inlineXML>
  </contentSet>
</newsItem>
```

### 4.4.1 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:20091107: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:

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

#### 4.4.1.1 Schema Reference:

```
xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
NewsML-G2_2.4-spec-NewsItem-Core.xsd"
```

This locates the IPTC's schema for NewsML-G2 version 2.2 on the local file system and associates it with the IPTC namespace.



Implementers need to consider whether there is a real need to validate 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.

#### 4.4.1.2 G2 standard and version:

```
standard="NewsML-G2" standardversion="2.4"
```

The Standard and Standard Version values **MUST** be aligned with the IPTC G2 Standards Schema being used.

The **ONLY** allowed values for @standard are either "NewsML-G2" or "EventsML-G2"; Depending on the version of the News Architecture being used, the following @standard/@standardversion pairs are allowed:

NAR version	@standard	@standardversion
1.1	NewsML-G2	2.0
	EventsML-G2	1.0
1.2	NewsML-G2	2.1
1.3	NewsML-G2	2.2
	EventsML-G2	1.1
1.4	NewsML-G2	2.3
	EventsML-G2	1.2
1.5	NewsML-G2	2.4
	EventsML-G2	1.3

The values of these attributes enable a G2 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.

#### 4.4.1.3 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 "power"

```
conformance="power"
```

#### 4.4.1.4 Catalog reference

```
<catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_11.xml"/>
```

This identifies the IPTC NewsCodes catalog, and will enable the G2 processor to de-reference QCodes as required. Any other catalogs used by the document must be declared in this section of the G2 document.

The IPTC recommends that a G2 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 G2 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.

#### 4.4.2 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 G2 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

```
<versionCreated>  
  2009-11-07T12:12:12+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 to be used by most public providers of G2, 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. English 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](#).

#### 4.4.3 Content Metadata `<contentMeta>`

This is the information directly associated with the content of the G2 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 G2 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 and time zone. The property is a TruncatedDateTime data type, which allows the value to be truncated from the right to a minimum of YYYY.

```
<contentCreated>2009-11-06</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 simply using the accepted geographical name “Washington”.<sup>3</sup>

```
<located literal="Washington" />
```

In later examples we will 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="LLeMeur">
  <name>Laurent Le Meur</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, cul ture 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.

The following line illustrates:

```
<subject type="cpnat: organi sation" 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 v2.4 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.

#### 4.4.4 The Content `<contentSet>`

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

```
<inlineXML contenttype="appl ication/xhtml+xml">
```

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

<sup>3</sup> Strictly, this value is an identifier, not a human-readable label. The 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.





## 5 Text

### 5.1 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 will show some of the essential building blocks of G2 in more detail, again using a piece of text content as an example. In subsequent chapters, we will cover the specific use cases of pictures, video and audio.

### 5.2 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-Paulson	
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	<a href="#">Henry Paulson</a>	
Organisations	<a href="#">US Treasury</a>	
Headline	Paulson: Must preserve bailout funds	
Byline	By Meredith Jameson	
Location / Date	Washington	21/11/2008
Body Text	<p>Treasury Secretary Henry Paulson on Tuesday said the unpredictable nature of the current financial crisis meant it was necessary to ensure that financial bailout money was not diverted to other uses.</p> <p>In testimony prepared for delivery to the U.S. House Financial Services committee, Paulson said the \$700-billion Troubled Asset Relief Program, or TARP, was intended to shore up the financial system and said there were other efforts under way to help homeowners avoid preventable foreclosures.</p>	

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.

### 5.3 The top level

#### 5.3.1 <newsItem>

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

##### 5.3.1.1 Standard

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

| standard="NewsML-G2"

##### 5.3.1.2 Standard Version

A string denoting the major and minor version of the G2 Standard being used:

| standardversion="2.2"

### 5.3.1.3 GUID

Each G2 document **MUST** provide an identifier that is guaranteed to be globally unique for all time and independent of location, (G2 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 G2 Items using a specification in RFC 3085bis. The syntax for a GUID using this scheme<sup>4</sup> 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-FI NANCE-PAULSON"
```

Any ID that can be guaranteed to be globally unique is valid for NewsML-G2. Some IPTC members 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 G2 which accurately convey a provider's intentions.

### 5.3.1.4 Further Attributes

**Conformance.** All 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 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 G2 document that has the same GUID, the version number **MUST** be incremented, not necessarily by 1.

```
| versi on="1"
```

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

```
| xml : l ang="en-US"
```

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

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

Namespace and schema information should also be added as appropriate. (see [The top level](#) above)

## 5.3.2 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:

---

<sup>4</sup>RFC 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 versi on attribute of <newsItem>



**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 Paulson. We have faked a link in the editing system which would point to some resource that can provide more information about Henry Paulson – 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_6.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="nature" 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](#).

### 5.3.3 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](#).

### 5.3.4 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:20081125T1205:US-FINANCE-PAULSON"
  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/
    NewsML-G2_2.2-spec-NewsItem-Core.xsd"
  standard="NewsML-G2"
  standardversion="2.2"
  xml:lang="en-US"
  >
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <catalogRef href="http://www.acmenews.com/synd/catalogs/anmcodes.xml" />
  <rightsInfo>
    <copyrightHolder literal="Acme News and Media LLC" />
  </rightsInfo>
```

## 5.4 Item Metadata

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



### 5.4.1 <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" />
```

### 5.4.2 <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" />
```

### 5.4.3 <versionCreated>

This contains the date and time 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>2008-11-21T16:25:32-05:00</versionCreated>
```

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

### 5.4.4 Optional Management Metadata elements

Element	Element Name	Datatype		Comment
Date Item First Created	firstCreated	DateTime		
Embargo	embargoed	Flexible		See <a href="#">17.4 Embargo</a>
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 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.

### 5.4.5 Links

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 semantic “Translated From” An example is given in [7.2.1.9](#)

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](#))

### 5.4.6 Summary

Our complete <itemMeta> section is:

```
<i temMeta>
  <i temCl ass qcode="ni nat: text" />
  <provi der l i t e r a l ="ANM" />
  <versi onCreated>
    2008-11-21T16: 25: 32-05: 00
  </versi onCreated>
  <pubStatus qcode="stat: usabl e" />
  <servi ce qcode="svc: USN">
    <name>U. S. News</name>
  </servi ce>
</i temMeta>
```

## 5.5 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.

### 5.5.1.1 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 G2 as Truncated DateTime data type. (See [Content Metadata](#)) 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>2008-11-21T15: 21: 06-05: 00</contentCreated>
<contentModi fi ed>2008-11-21T16: 22: 45-05: 00</contentModi fi ed>
```

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

```
<l o c a t e d l i t e r a l =" 20001">
  <name>Washi ngton D. C. </name>
</l o c a t e d>
```

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 human-readable <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>Meredit h Jameson</name>
</creator>
```

#### 5.5.1.2 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 “lift” metadata that is 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 human-readable name of the Category using the <name> tag:

```
<subject qcode="ANMCat: TRS">
  <name>Treasury News</name>
</subject>
<subject qcode="ANMCat: FIN">
  <name>Fi nanci al 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: organi sati on" qcode="ANMorg: 5768933">
  <name>U. S. Treasury Department</name>
</subject>
<subject type="cpnat: person" qcode="ANMpers: 9999999">
  <name>Hank Paul son, 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-Paul son</slugline>
```

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

```
<headline>Paul son: Must preserve bai l out funds</headline>
```

## 5.6 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>
            <h1> Paulson: Must preserve bailout funds</h1>
          </headline>
          <byline>
            <byttl>By Meredith Jameson</byttl>
          </byline>
        </body.head>
        <body.content>
          <p>
            Treasury Secretary Henry Paulson on Tuesday said the unpredictable nature of the
            current financial crisis meant it was necessary to ensure that financial bailout
            money was not diverted to other uses.
          </p>
          <p>
            In testimony prepared for delivery to the U.S. House Financial Services committee,
            Paulson said the $700-billion Troubled Asset Relief Program, or TARP, was intended
            to shore up the financial system and said there were other efforts under way to
            help homeowners avoid preventable foreclosures.
          </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.

## 5.7 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:20081125T1205:US-FINANCE-PAULSON"
  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/
    NewsML-G2_2-spec-NewsItem-Core.xsd"
  standard="NewsML-G2"
  standardversion="2.2"
  xml:lang="en-US"
>
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <catalogRef href="http://www.acmenews.com/synd/catalog/anmcodes.xml" />
  <rightsInfo>
    <copyrightHolder literal="Acme News and Media LLC" />
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="ni:at:text" />
    <provider literal="ANM" />
    <versionCreated>2008-11-25T16:25:32-05:00</versionCreated>
    <pubStatus qcode="stat:usable" />
  </itemMeta>
  <contentMeta>
    <contentCreated>2008-11-25</contentCreated>
    <located literal="Washington DC" />
    <creator literal="Meredith Jameson" />
    <language tag="en-US" />
    <subject qcode="ANMCat:TRS">
      <name>Treasury News</name>
    </subject>
    <subject qcode="ANMCat:FIN">
      <name>Financial News</name>
    </subject>
  </contentMeta>
```



```
</subject>
<subject type="cpnat:organization" qcode="org:5768933">
  <name>U. S. Treasury Department</name>
</subject>
<subject type="cpnat:person" qcode="pers:9999999">
  <name>Hank Paulson, Treasury Secretary</name>
</subject>
<slugline>US-Finance-Paulson</slugline>
<headline>Paulson: 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> Paulson: Must preserve bailout funds</hl1>
          </headline>
          <byline>
            <byttl>By Meredith Jameson</byttl>
          </byline>
        </body.head>
        <body.content>
          <p>Treasury Secretary Henry Paulson on Tuesday said the
            unpredictable nature of the current financial crisis meant it
            was necessary to ensure that financial bailout money was not
            diverted to other uses.</p>
          <p>In testimony prepared for delivery to the U.S. House
            Financial Services committee, Paulson said the $700-billion
            Troubled Asset Relief Program, or TARP, was intended to shore
            up the financial system and said there were other efforts under
            way to help homeowners avoid preventable foreclosures.</p>
        </body.content>
      </body>
    </nitf>
  </inlineXML>
</contentSet>
</newsItem>
```

## 5.8 Other content options

### 5.8.1 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 the G2 standards 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.

### 5.8.2 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" />
  Treasury Secretary Henry Paulson on Tuesday said the unpredictable nature of the
  current financial crisis meant it was necessary to ensure that financial bailout
  money was not diverted to other uses.
  In testimony prepared for delivery to the U.S. House Financial Services committee,
  Paulson said the $700-billion Troubled Asset Relief Program, or TARP, was intended to
  shore up the financial system and said there were other efforts under way to help
  homeowners avoid preventable foreclosures.
</inlineData>
</contentSet>
```





It is also possible for <inlineData> to hold plain text or 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

## 6 Pictures and Graphics

### 6.1 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 some of the metadata used in the example are courtesy of Thomson Reuters.

*Note that the sample code is NOT a guide to receiving NewsML-G2 from Thomson Reuters.*

#### 6.1.1 Use Case



A picture of Brad Pitt and Angelina Jolie attending the premiere of his latest film *The Curious Case of Benjamin Button*, will be provided to customers in three sizes: High Resolution (the original resolution using the U.S. SWOP Coated v2 colour space), Web (a lower resolution RGB image for Web use), and Thumbnail (a small RGB image intended only to give an idea of the image content). These three images are termed different **renditions** of the same image. We will convey this content and its metadata in a single NewsML-G2 document.

### 6.2 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](http://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 one of the Adobe Photoshop CS2 File Info screens.

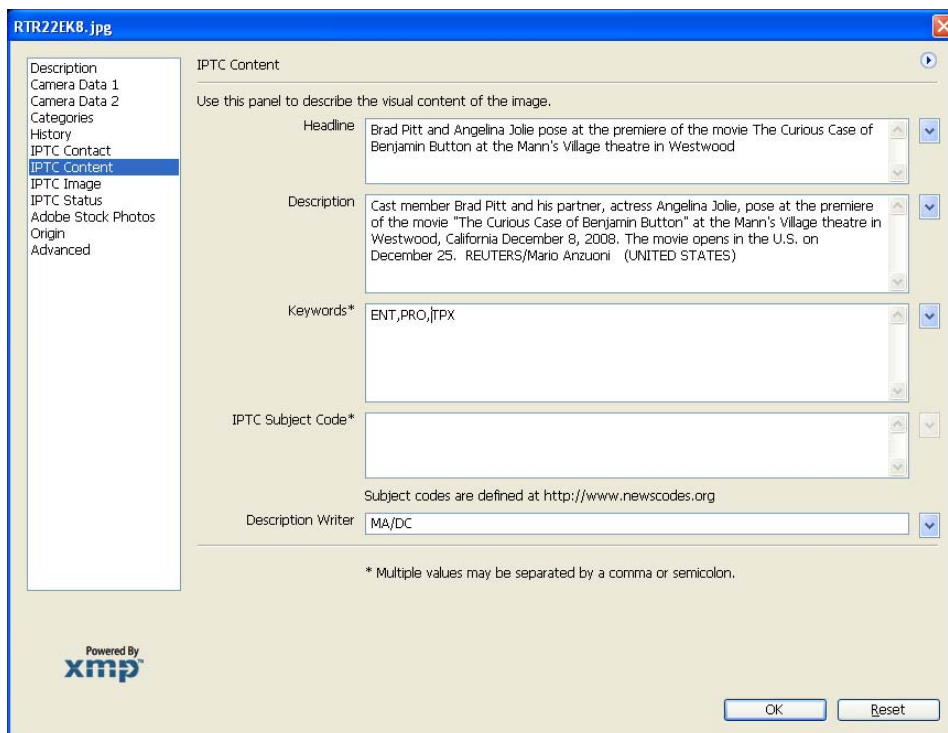


Figure 5: One of the IPTC Custom metadata panels available in Adobe Photoshop CS



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.

#### 6.2.1.1 Reconciling photo metadata

A situation may arise where the metadata expressed in the 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 G2 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 G2 <located> and <contentCreated> properties exist in the embedded metadata, these MAY take precedence over the G2 metadata, because they may have been created by technical means, but this is subject to guidance from the provider.

## 6.2.2 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 the G2 Standards, so we need to set the appropriate conformance level to “Power” in the <newsItem> element.

#### 6.2.2.1 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
  standard="NewsML-G2"
  standardversion="2.2"
  conformance="power"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="tag:example.com,2008:ART-ENT-SRV:USA20081220098658" version="1"
>
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml"
  />
  <!-- also using provider-specific (albeit fictitious) taxonomies -->
  <catalogRef
    href="http://www.example.com/customer/cv/catalog4customers-1.xml"
  />
```

#### 6.2.2.2 Rights

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



```
<rightsInfo>
  <copyrightHolder literal="RTRS" />
  <copyrightNotice>(c) Copyright Thomson Reuters 2008. Click For
    Restrictions - http://about.reuters.com/fullLegal.asp
  </copyrightNotice>
  <usageTerms xml:lang="en">NO ARCHIVAL OR WIRELESS USE</usageTerms>
</rightsInfo>
```

### 6.2.2.3 Provider

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

```
<provider literal="reuters.com"/>
```

### 6.2.2.4 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 in G2. 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 G2 property that most closely aligns with the provider's intention.

In the example, the provider intends the contents of the Title field to carry the Slugline that journalists use to identify the picture in their content management system. The contents the Title field can therefore be mapped to the G2 <slugline> property. In cases like this, an implementer could choose to map only to <slugline>, or to map to both <title> and <slugline>. Here, the Title field is mapped to <title> in <itemMeta> and copied into the <slugline> property in <contentMeta>.

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 Status<sup>5</sup>.

```
<itemMeta>
  <itemClass qcode="ni nat: picture" />
  <provider literal="reuters.com"/>
  <versionCreated>
    2008-12-20T13:20:00Z
  </versionCreated>
  <firstCreated>
    2008-12-20T13:20:00Z
  </firstCreated>
  <publishStatus qcode="stat: usable" />
  <title>US-FILM-BUTTON</title>
</itemMeta>
<contentMeta>
  . . . .
  <slugline>US-FILM-BUTTON</slugline>
  . . . .
</contentMeta>
```

### 6.2.2.5 City, State/Province, Country

G2 provides a <located> element in the <contentMeta> block to describe 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 at an office elsewhere from the event)
- ❖ 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](#))

<sup>5</sup> Note that these dates in <itemMeta> refer to the News Item, NOT the content. The date stamps for the content are placed in <contentMeta>



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 event took place in the city of Westwood, California. We may use two child elements of <subject>:

- ❖ <name> allows us to literally name the place using plain text, and
- ❖ <broader><sup>6</sup> allows us to convey the semantic of Westwood as part of the broader entity of California, which is part of a broader entity of the United States.

The <subject> and <broader> elements are flexible – their attributes can be:

- ❖ @literal (place name in plain text) 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 a geographical description. For example

```
<!-- @literal is a zip code -->
<subject type="cptype: zip" literal="96137">
  <name xml:lang="en">Westwood, California, USA</name>
</subject>
```

Our example will show @type and @qcode, using some fictitious schemes:

- ❖ Concept Type (cptype:) will allow us to declare that the entity identified by @qcode is either a city, state/province, or country, thus:
  - | type="cptype: city"
- ❖ City (city:) allows us to uniquely identify the city of Westwood from the controlled vocabulary of the world’s cities: Note that there is NO correspondence between “city:” in @type and “city:” in @qcode.
  - | qcode="city: 28398"
- ❖ State/Province and Country will also use their respective scheme aliases, “state:” and “country:”

The completed <subject> structure will be:

```
<subject type="cptype: city" qcode="city: 28398">
  <name>Westwood</name>
  <broader type="cptype: statprov" qcode="state: 3959">
    <name>California</name>
  </broader>
  <broader type="cptype: country" qcode="country: US">
    <name>United States</name>
  </broader>
</subject>
```

#### 6.2.2.6 Creator

We will use the <creator> element, which also has flexible properties. We will use a literal value:

```
<creator literal="MANzuoni" />
```

#### 6.2.2.7 Source, Transmission Reference

In our example, both of these values in the picture metadata are, in effect, alternative identifiers for the picture that may be needed by customer systems. At PCL, we may use the <altId> element to convey this information. <altId> may optionally be refined using a QCode to describe the context. We will use a (fictitious) scheme alias of “idtype:” and the two values we will use are “systemRef” for Source and “transmitRef” for the Transmission Reference. 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 G2 processor to identify or locate the resource.

```
<altId type="idtype: systemRef">X90045</altId>
<altId type="idtype: transmitRef">MA203</altId>
```

<sup>6</sup> <broader> is only available at Power Conformance Level, which is why we set @conformance to “power” in <newsItem>



#### 6.2.2.8 IPTC Subject Code

The IPTC maintains a three-level scheme known as Subject NewsCodes for classifying content according to its subject matter (see NewsCodes at [www.iptc.org](http://www.iptc.org) or visit [www.newscodes.org](http://www.newscodes.org)). As discussed previously, the subject matter of content uses the <subject> element and at PCL we may use <broadener> to refine the description. We will also use @qcode and @type using the appropriate NewsCodes. A subject code is an “abstract” concept (as opposed to an entity such as “person”) The recommended scheme alias for the Subject NewsCodes is “subj:” and alias for the “nature of the concept” NewsCodes scheme is “cpnat:”

The picture of Brad and Angelina fits neatly into the “Cinema” category, which is part of Arts, Culture and Entertainment section of the Subject NewsCodes. This equates to a numeric code of 01005000. Using PCL and the <broadener> element, we can express this as:

```
<subject type="cpnat:abstract" qcode="subj:01005000">
  <name xml:lang="en">Cinema</name>
  <name xml:lang="fr">Cinéma</name>
  <broadener type="cpnat:abstract" qcode="01000000">
    <name xml:lang="en">Arts, Culture and Entertainment</name>
    <name xml:lang="fr">Arts, culture, et spectacles</name>
  </broadener>
</subject>
```

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

#### 6.2.2.9 Headline, Description

These two XMP fields map directly to elements of the same name in 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 markup and features such as @ruby for supporting Asian character sets.

```
<headline xml:lang="en">Brad Pitt and Angelina Jolie pose at the premiere of the
  movie The Curious Case of Benjamin Button at the Mann's Village theatre in
  Westwood
</headline>
<description xml:lang="en" role="drol:caption">Cast member Brad Pitt and his partner,
  actress Angelina Jolie, pose at the premiere of the movie "The Curious Case
  of Benjamin Button" at the Mann's Village theatre in Westwood,
  California December 8, 2008. The movie opens in the U.S. on December 25.
  REUTERS/Mario Anzuoni (UNITED STATES)
</description>
```

For a more detailed description of mapping photo metadata to G2, see [Mapping Embedded Photo Metadata to G2](#)

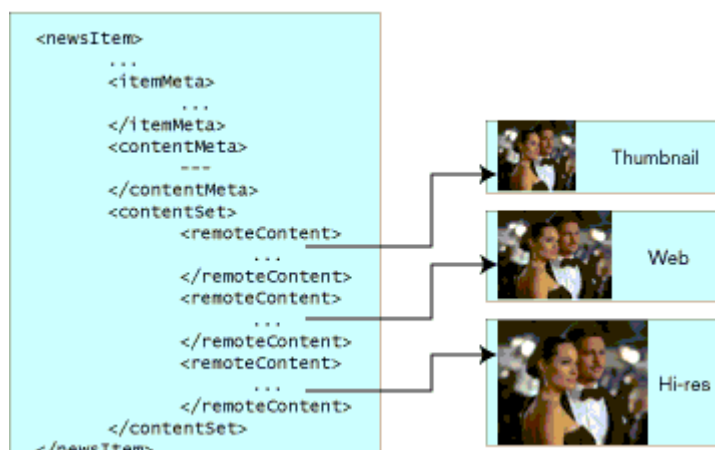
### 6.3 Photo data

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 G2 Standards permit 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.

#### 6.3.1 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.





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.

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

## 6.3.2 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.

### 6.3.2.1 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 Hi-res. We may need to assert different rights to the each of the referenced resources. The Hi-res version 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="some-i d" />
```

### 6.3.2.2 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 alias of “rnd:” and contains (amongst others) the values that we need: hiRes, web, thumbnail. Thus using the appropriate NewsCode, the high resolution version of the picture may be identified as:

```
<remoteContent id="some-i d" rendition="rnd: hi Res" />
```

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 “hi-res” images conveyed in the same Item.

## 6.3.3 Target Resource Attributes

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



### 6.3.3.1 Hyperlink (@href)

An IRI, e.g.

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

Or (amongst other possibilities)

```
<remoteContent href="file:///some/path/foo.jpg" />
```

### 6.3.3.2 Resource Identifier Reference (@residref)

An XML Schema string e.g.

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

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

### 6.3.3.3 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
  href="http://example.com/2008-12-20/pictures/foo.jpg"
  residref="tag:example.com, 2008:PIX:F0020081220098658"
  version="1"
/>
```

### 6.3.3.4 Content Type

The MIME type of the target resource

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

### 6.3.3.5 Size

Indicates the size of the target resource in bytes.

```
size="3764418"
```

## 6.3.4 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.

### 6.3.4.1 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



```
wi dth="2500" wi dthuni t="di mensi onuni t: pi xel s"
hei ght="2075" hei ghtuni t="di mensi onuni t: pi xel s"
```

#### 6.3.4.2 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.

```
ori entati on="1"
```

#### 6.3.4.3 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.

```
col ourspace="col sp: USSW0Pv2"
```

#### 6.3.4.4 Resolution

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

```
resol uti on="300"
```

### 6.4 The Complete Picture

The **LISTING 3** below is given for completeness. Some of the optional information such as @version and @resolution has been omitted. The images are identified/located using @href alone.

#### LISTING 3 Conveying a Picture in NewsML-G2

```
<?xml versi on="1. 0" encodi ng="UTF-8"?>
<newsI tem
  gui d="tag: exampl e. com, 2008: ART-ENT-SRV: USA20081220098658"
  versi on="1"
  xml ns="http: //i pt c. org/std/nar/2006-10-01/"
  xml ns: xsi ="http: //www. w3. org/2001/XMLSchema-i nstance"
  xsi : schemaLocati on="http: //i pt c. org/std/nar/2006-10-01/
    NewsML-G2_2. 2-spec-NewsI tem-Power. xsd"
  standard="NewsML-G2"
  standardversi on="2. 2"
  conformance="power" >
  <catalogRef
    href="http: //www. i pt c. org/std/catalog/catalog. IPTC-G2-Standards_6. xml " />
  <catalogRef
    href="http: //www. exampl e. com/customer/cv/catalog4customers-1. xml " />
  <ri ghtsI nfo>
    <copyri ghtHol der l i t e r a l ="RTRS" />
    <copyri ghtNoti ce>(c) Copyright Thomson Reuters 2008. Click For
      Restrictions - http: //about. reuters. com/full l e g a l . a s p
    </copyri ghtNoti ce>
    <usageTerms xml : l ang="en">MUST COURTESY PARAMOUNT PICTURES
      FOR USE OF "THE CURIOUS CASE OF BENJAMIN BUTTON" WITH NO ARCHIVAL OR
      WI RELESS USE</usageTerms>
  </ri ghtsI nfo>
  <i temMeta>
    <i temCl ass qcode="ni nat: pi ctur e" />
    <provi der l i t e r a l ="reuters. com" />
    <versi onCreated>2008-12-20T13: 20: 00Z</versi onCreated>
    <fi rstCreated>2008-12-19T23: 25: 35-08: 00</fi rstCreated>
    <pubStatus qcode="stat: usable" />
    <ti t l e>US-FI LM-BUTTON</ti t l e>
  </i temMeta>
  <contentMeta>
    <contentCreated> 2008-12-19T23: 04: 00-08: 00</contentCreated>
    <contentModi fi ed> 2008-12-20T13: 15: 00Z</contentModi fi ed>
    <creator l i t e r a l ="MANzuoni " />
    <al t i d type="i dtype: systemRef">X90045</al t i d>
    <al t i d type="i dtype: transmi tRef">MA203</al t i d>
    <l anguage tag="en" rol e="l rol : voi ceOver" />
    <subj ect type="cptype: ci ty" qcode="ci ty: 28398">
      <name>Westwood</name>
      <broad er type="cptype: statprov" qcode="stat: 3959">
```



```

    <name>Cal i forni a</name>
  </broader>
  <broader type="cptype: country" qcode="country: US">
    <name>Uni ted States</name>
  </broader>
</subject>
<subject type="cpnat: abstract" qcode="subj: 01005000">
  <name xml:lang="en">Ci nema</name>
  <name xml:lang="fr">Ci néma</name>
  <broader type="cpnat: abstract" qcode="subj: 01000000">
    <name xml:lang="en">Arts, Cul ture and Entertai nment</name>
    <name xml:lang="fr">Arts, cul ture, et spectacl es</name>
  </broader>
</subject>
<slugline> US-FILM-BUTTON</slugline>
<headline xml:lang="en">Brad Pitt and Angelina Jolie pose at
  the premiere of the movie The Curious Case of Benjamin Button at the
  Mann&apos;s Village theatre in Westwood</headline>
<description xml:lang="en" role="drol: caption">Cast member Brad Pitt
  and his partner, actress Angelina Jolie, pose at the premiere of the
  movie &quot;The Curious Case of
  Benjamin Button&quot; at the Mann&apos;s Village theatre in Westwood,
  California December 8, 2008. The movie opens in the U.S. on December 25.
  REUTERS/Mario Anzuoni (UNITED STATES)
</description>
</contentMeta>
<contentSet>
  <remoteContent
    href="http://www.example.com/content/pictures/hi res/20081220/US-FILM-
    BRAD_001_HI.jpg"
    rendition="rnd: hi Res"
    size="3764418"
    contenttype="image/jpeg"
    width="2500"
    height="2075"
    colospace="col sp: USSWOPv2" />
  <remoteContent
    href="http://www.example.com/content/pictures/web/20081220/US-FILM-
    BRAD_001_WEB.jpg"
    rendition="rnd: web"
    size="11134"
    contenttype="image/jpeg"
    width="375"
    height="311"
    colospace="col sp: AdobeRGB" />
  <remoteContent
    href="http://www.example.com/content/pictures/thumbnai ls/20081220/US-FILM-
    BRAD_001_THUMB.jpg"
    rendition="rnd: thumbnai l "
    size="2761"
    contenttype="image/jpeg"
    width="100"
    height="83"
    colospace="col sp: AdobeRGB" />
</contentSet>
</newsItem>

```



This page intentionally blank



## 7 Audio and Video

### 7.1 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.

The G2 Standards are designed to allow all organisations, whether traditional broadcasters or not, to access and exchange audio and video in a professional workflow. Existing G2 features and extension points enable proprietary formats to be “mapped” to G2 to achieve freedom of exchange amongst a wider circle of information partners.

Audio and video have an additional dimension – duration of time – which is not present in text and pictures, so we may 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.<sup>7</sup> 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, 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.

### 7.2 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 humorist 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.

#### 7.2.1 Metadata

##### 7.2.1.1 Provider

G2 enables us to give 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">
          <name xml:lang="en">Switzerland</name>
        </country>
      </address>
    </contactInfo>
  </organisationDetails>
</provider>
```

<sup>7</sup> Note that this complies with the basic 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 put together by a single reporter or editor from several different reports, ,



```

        </country>
    </address>
</contactInfo>
</organisationDetails>
</provider>

```

#### 7.2.1.2 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>

```

#### 7.2.1.3 Editorial Note

A human-readable note for receivers in <itemMeta>

```

<edNote>Originally broadcast in Germany</edNote>

```

#### 7.2.1.4 Located

We use the <located> and <broadcaster> 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>
  <broadcaster type="cptype: statprov" qcode="state: 2365">
    <name>Berlin</name>
  </broadcaster>
  <broadcaster type="cptype: country" qcode="country: DE">
    <name>Germany</name>
  </broadcaster>
</located>

```

#### 7.2.1.5 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: Technical Origin">
  <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 human-readable 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.

#### 7.2.1.6 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.

#### 7.2.1.7 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: bi og">
  <name xml:lang="en">Bi ography</name>
  <name xml:lang="fr">bi ographi e</name>
</genre>
```

The subject is "Arts, Culture and Entertainment", and a narrower definition would be "Animation". In **IPTC Subject Code**, we showed the use of <broadener> to refine the subject "Cinema" as part of "Arts, Culture and Entertainment", in this case, we will invert the relationship using <narrower>

```
<subject type="cpnat: abstract" qcode="subj: 01000000">
  <name xml:lang="en">Arts, Cul ture and Entertainment</name>
  <name xml:lang="fr">Arts, cul ture, et spectacles</name>
  <narrower type="cpnat: abstract" qcode="01025000">
    <name xml:lang="en">Ani mati on</name>
    <name xml:lang="fr">Dessi n ani mé</name>
  </narrower>
</subject>
```

#### 7.2.1.8 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 exhibi tion
opened in Berlin in honour of German humorist Vicco von Bü low, 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 Bü low
entering exhibi tion<br/>- vs. Loriot and media<br/>-sot Vicco von Bü low <br/>"Since 85
years I didn't succeed in pursuing a job that could be called a profession."<br/>- vs
exhibi tion<br/>- sot Irm Herrmann, actress<br/>"Loriot is timeless. You always can
watch him and I can always laugh."<br/>-actor Ulrich Matthes in exhibi tion<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 `<br />`

#### 7.2.1.9 Links

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

- ❖ assert relationships to other G2 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/thi_s_item/index.html">
```

#### 7.2.1.10 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>Vico von Bülow entering exhibition</description>  
</partMeta>
```



## 7.2.2 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:

### 7.2.2.1 Duration (@duration)

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

### 7.2.2.2 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 **7.2.1.10**)

### 7.2.2.3 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”

### 7.2.2.4 Video Frame Rate (@videoframerate)

An integer value indicating the rate, in frames per second, that the video should be played out in order to achieve the correct visual effect

### 7.2.2.5 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"/>
```

## 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/
    NewsML-G2_2.3-spec-NewsItem-Power.xsd"
  standardversion="2.3"
  guid="tag:example.com, 2008: 407490" version="1" standard="NewsML-G2"
  conformance="power" xml:lang="en">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_11.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 nat: video" />
<provider qcode="providercode: EBU">
  <name>European Broadcasting Union - EVN</name>
  <organizationDetails>
    <contactInfo>
      <web>http://www.eurovision.net</web>
      <phone>+41 22 717 2869</phone>
      <email>features@eurovision.net</email>
      <address role="AddressType: Office">
        <line>Eurovision Sports News Exchanges</line>
        <line>L Ancienne Route 17 A</line>
        <line>CH-1218</line>
        <locality literal="Grand-Saconnex" />
        <country qcode="ISOCountryCode: ch">
          <name xml:lang="en">Switzerland</name>
        </country>
      </address>
    </contactInfo>
  </organizationDetails>
</provider>
<versionCreated>2008-11-07T10:54:04Z</versionCreated>
<firstCreated>2008-11-05T15:22:28Z</firstCreated>
<pubStatus qcode="stat: usable" />
<service qcode="servicecode: EUROVISION">
  <name>Eurovision services</name>
</service>
<edNote>Originally broadcast in Germany</edNote>
<link rel="itemrel: preview"
  href="http://www.example.com/video/2008-12-22/20081222-PNN-1517-407624/index.html" />
</itemMeta>
<contentMeta>
  <contentCreated> 2008-12-22T23:04:00-08:00</contentCreated>
  <located type="cptype: city" qcode="city: 345678">
    <name>Berlin</name>
    <broadcaster type="cptype: statprov" qcode="state: 2365">
      <name>Berlin</name>
    </broadcaster>
    <broadcaster type="cptype: country" qcode="country: DE">
      <name>Germany</name>
    </broadcaster>
  </located>
  <creator qcode="codesource: DEZDF">
    <name>Zweites Deutsches Fernsehen</name>
    <organizationDetails>
      <location literal="MAINZ" />
    </organizationDetails>
  </creator>
  <contributor qcode="codeorigin: DEZDF" role="rolecode: Technical Origin">
    <name>Zweites Deutsches Fernsehen</name>
  </contributor>
  <creator qcode="codesource: GBRTV">
    <name>Reuters Television Ltd</name>
  </creator>
  <language tag="en" role="languagecode: PartNarrated">
    <name>English</name>
  </language>
  <genre qcode="genre: biog">
    <name xml:lang="en">Biography</name>
    <name xml:lang="fr">biographie</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">Animation</name>
      <name xml:lang="fr">Dessin animé</name>
    </narrower>
  </subject>
</headline>Loriot retrospective</headline>
<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 Hamann 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.
    </description>
    <description role="descrole:shotlist">
        Berlin, 21/12/2008
    </description>
    - vs. Vicco von Bülow entering exhibition
    </description>
    - vs. Loriot and media
    </description>
    - sot Vicco von Bülow
    </description>
    "Since 85 years I didn't succeed in pursuing a job that could be
    called a profession."
    </description>
    - vs exhibition
    </description>
    - sot Irm Herrmann, actress
    </description>
    "Loriot is timeless. You always can watch him and I can always
    laugh."
    </description>
    - actor Ulrich Matthes in exhibition
    </description>
    sot Ulrich Matthes, actor
    </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>
</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>Loriot 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:Natural Sound"/>
    <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="languagecode:VoiceOver"/>
  <description>"I would say: one of the great German classics. Goethe, Kleist, Schiller, Thomas Mann, Lortie. 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="vcodec:c008"
    video framerate="25"
    video aspectratio="16:9" />
</contentSet>
</newsItem>
```

### 7.3 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 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:

#### 7.3.1.1 Audio Bit Rate (@audiobitrate)

A positive integer indicating kilobits per second (Kbps)

#### 7.3.1.2 Audio Sample Rate (@audiosamplerate)

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

#### 7.3.1.3 Video Average Bit Rate (@videoavgbitrate)

A positive integer indicating the average bit rate in Kbps) of a video encoded with a variable bit 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"
    video aspectratio="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"
    video aspectratio="16:9" />
</contentSet>
```



## 8 Packages

### 8.1 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.

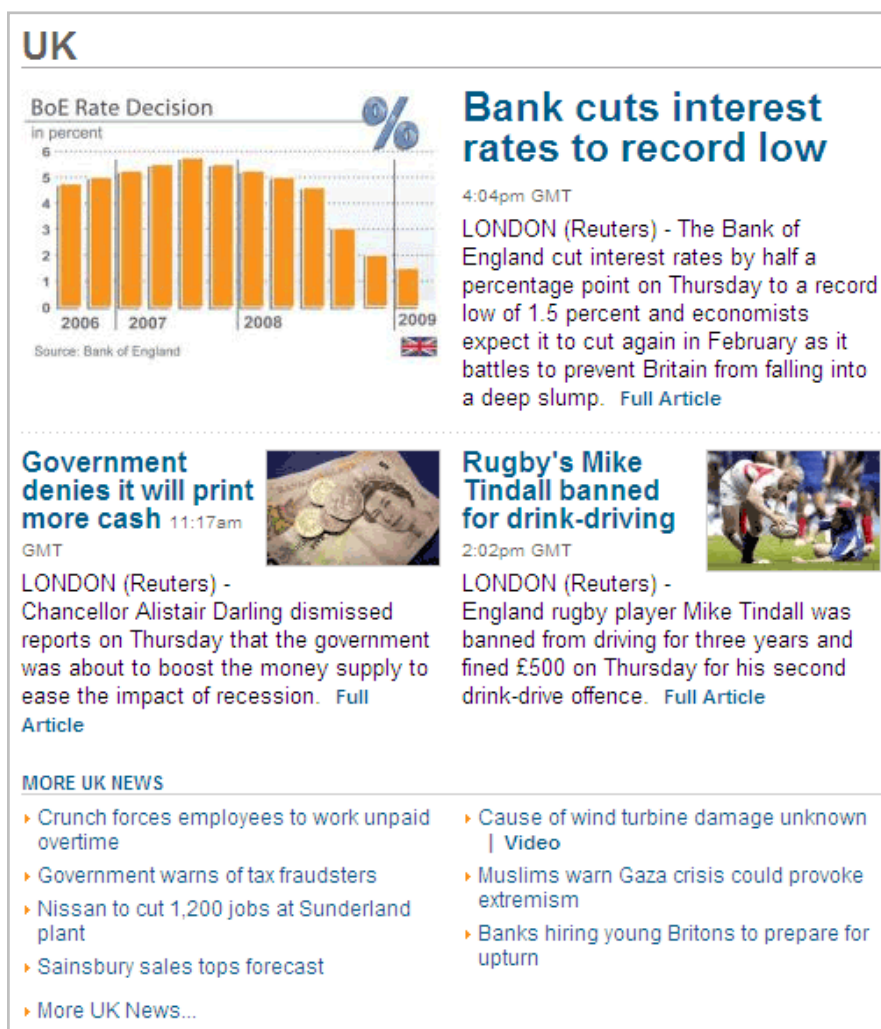


Figure 6: A "Top Ten" package on the Web

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.

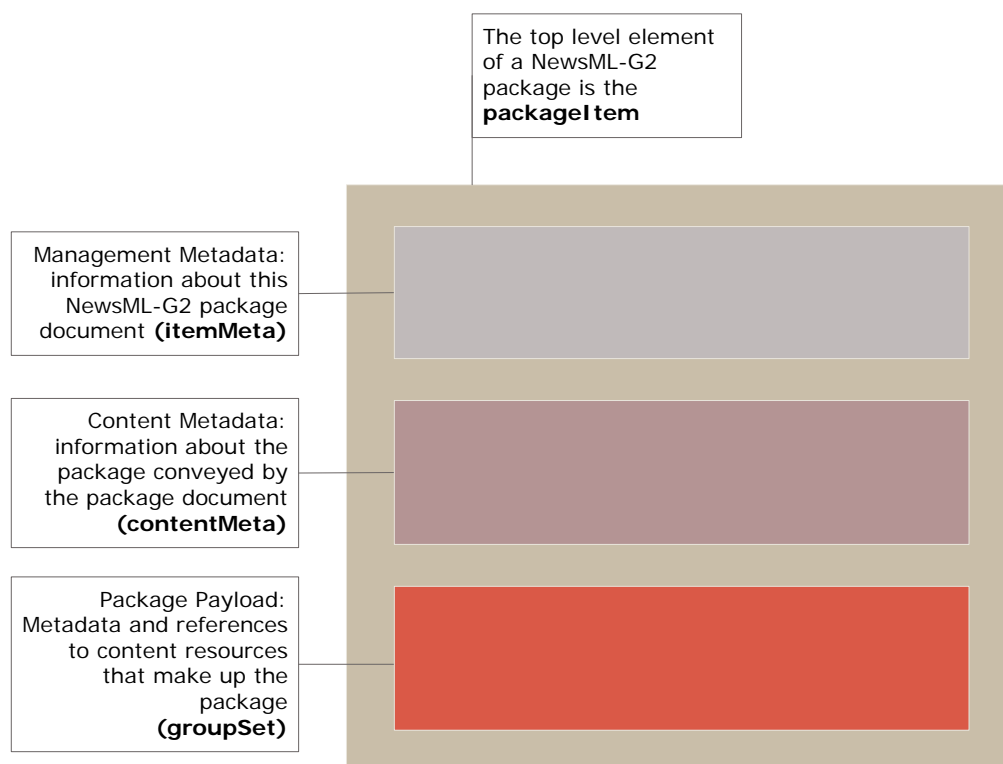
G2 is flexible in allowing a provider to package content that has already been published, or a package may be sent together with the content resources to which it refers (see [Exchanging News: News Messages](#)).

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 7: 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 either a G2 Item (upper-case “I”) or other discrete item of content. A G2 Item is specifically indicated by capitalizing the word “Item”



As a member of the G2 Family, Package Items share many properties with News Items; compare the model as shown in [Figure 7](#) to the News Item model shown in [Figure 4](#) and they are seen to be very 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 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.

## 8.2 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

## 8.3 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 8](#). (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 8: Simple Package using Item References <itemRef>

### 8.3.1 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 QCode value. Thus

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

### 8.3.2 Item Reference <itemRef>

The <itemRef> element identifies an item or a Web resource using the G2 LinkType template (CCL), or G2 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 G2 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 G2 News Item. The following are the allowed values for @contenttype referencing G2 Items:

G2 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

@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="300039">
      <title>Barack Obama arrive à Washington</title>
    </itemRef>
  </group>
</groupSet>
```

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

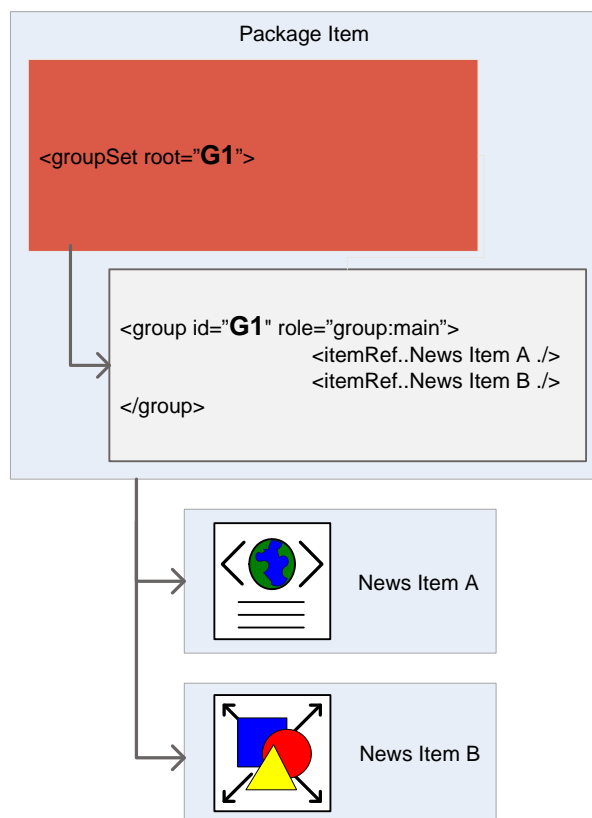


Figure 9: Simple Package Relationship

At PCL, the G2 Link1Type data type's Hint and Extension Point allows us to extract further properties from the target resource and place them as child elements of the Item Reference as 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 when using the Hint and Extension Point, the immediate child properties of <itemMeta> or <contentMeta>, can be used without the parent element; the extracted 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 resIdref="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 resIdref="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>
```



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

## 8.4 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 10](#) and [Figure 11](#)) and in code in [LISTING 8](#).

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 10: 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

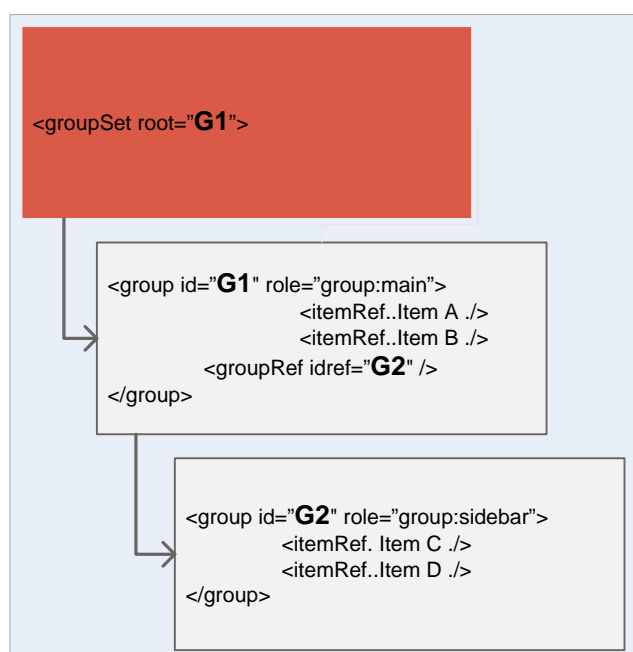


Figure 11: Relationship diagram of Hierarchical Package

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

#### LISTING 8 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>
  </group>
  <group id="G2" role="group:sidebar">
    <itemRef..Item C ./>
    <itemRef..Item D ./>
  </group>
</groupSet>

```

```
<groupRef idref="G2" />
</group>
<group id="G2" role="group: sidebar">
  <itemRef residref="urn: newsml : iptc. org: 20081007: tutorial -i tem-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 -i tem-D"
    contenttype="application/vnd. iptc. g2. newsitem+xml "
    size="350280">
    <title>Hillary Clinton à une rassemblement à New York</title>
  </itemRef>
</group>
</groupSet>
```

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

## 8.5 List Type Package Structure (ordered, bag, alternative)

**LISTING 8** 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-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 G2 document; we can freely re-use these ids in other G2 packages.

The relationship is shown in the diagrams (**Figure 12** and **Figure 13**).

```

<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 12: Using `<groupRef>` to create an unordered “bag” of groups of Items

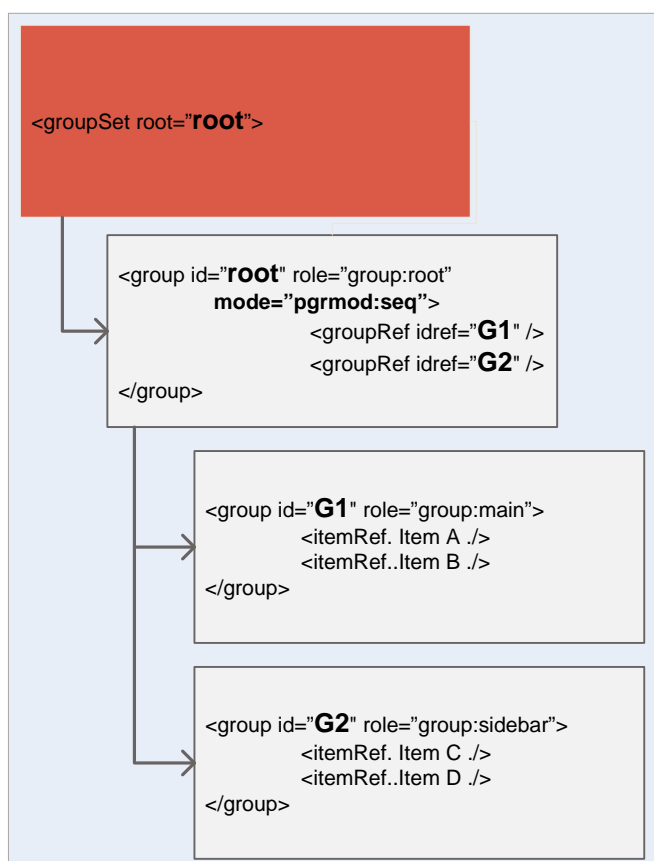


Figure 13: Groups G1 and G2 are both children of the Root Group





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

### LISTING 9 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: tutori al -i tem-A"
      contenttype="appli cati on/vnd. iptc. g2. newsi tem+xml "
      size="2345">
      <ti tle>Obama annonce son équipe</ti tle>
    </i temRef>
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutori al -i tem-B"
      contenttype="appli cati on/vnd. iptc. g2. newsi tem+xml "
      size="300039">
      <ti tle>Barack Obama arrive à Washington</ti tle>
    </i temRef>
  </group>
  <group id="G2" role="group: si debar">
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutori al -i tem-C"
      contenttype="appli cati on/vnd. iptc. g2. newsi tem+xml "
      size="1503">
      <ti tle>Clinton reprend son rôle de chef de la santé</ti tle>
    </i temRef>
    <itemRef residref="urn: newsml: iptc. org: 20081007: tutori al -i tem-D"
      contenttype="appli cati on/vnd. iptc. g2. newsi tem+xml "
      size="350280">
      <ti tle>Hillary Clinton à une rassemblement à New York</ti tle>
    </i temRef>
  </group>
</groupSet>
```

## 8.6 Package Processing Considerations

### 8.6.1 Other G2 Items

In the above examples, the referenced resources in the package have been G2 News Items, but `<itemRef>` may also refer to other G2 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:

```
<i temRef residref="tag: afp. com, 2008: TX-PAR: 20080529: JYC99"
  contenttype="appli cati on/vnd. iptc. g2. packagei tem+xml "
  size="28047">
  <i temCl ass qcode="ni nat: composi te" />
  <provi der l i t e r a l="AFP"/>
  <pubSt atus qcode="stat: usabl e"/>
  <ti tle>Tiger Woods cherche son retour</ti tle>
  <descri pti on 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.
  </descri pti on>
</i temRef>
```

## 8.6.2 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 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:

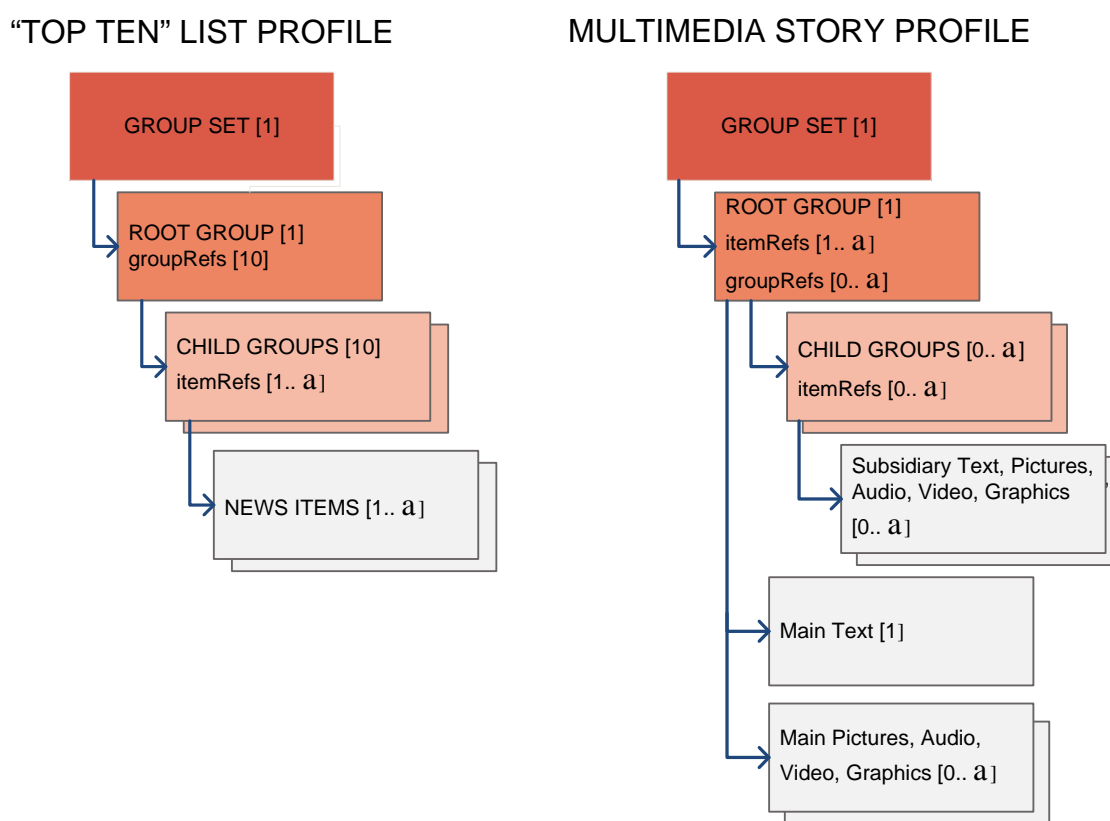


Figure 14: Diagrams of Package Profiles. The numbers in brackets indicate the required items

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 14**)

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 human-readable 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"
/>
```

### 8.6.3 A “Top Ten” Package Example

In the example in **LISTING 10**, we will show an ordered list of references to G2 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 14**. 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 human-readable <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 10 The complete Top Ten style News Package

```
<?xml version="1.0" encoding="UTF-8"?>
<packageItem
  standard="NewsML-G2"
  standardversion="2.2"
  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/
    NAR_1.3-spec-PackageItem-Power.xsd"
  guid="tag: example. com, 2008: UK-NEWS-TOPTEN: UK20081220098658" version="16">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <catalogRef
    href="http://www.example.com/customer/cv/catalog4customers-1.xml" />
  <itemMeta>
    <itemClass qcode="ni nat: composite" />
    <provider literal="MyNews" />
    <versionCreated> 2008-12-20T16: 40: 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: uktop">
      <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="inat: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="inat: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="inat: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="inat: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="inat: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>
  </group>

```

```

        </descri pti on>
    </i temRef>
    <i temRef resi dref="tag: afp. com, 2008: PX-PAR: 20080529: JYC51"
        contenttype="appl ication/vnd. iptc. g2. newsi tem+xml " si ze="31285">
        <i temCl ass qcode="ni nat: pi ctur e" />
        <provi der l i teral="MyNews" />
        <pubSt atus qcode="stat: usabl e" />
        <headl i ne>Engl and i nternati onal Mi ke Ti ndal l </headl i ne>
        <descri pti on rol e="drol : capti on">
            Engl and' s Mi ke Ti ndal l (L) scores hi s try despite a l ate
            tackl e from Li onel Beauxi s (R) of France dur i ng thei r Si x Nat i ons
            match at Twi ckenham March 11, 2007.
        </descri pti on>
    </i temRef>
</group>
<group i d="G4" rol e="group: packag e" />
<!-- more groups i d="G5" to i d="G9" -->
<group i d="G10" rol e="group: packag e" />
</groupSet>
</packagel tem>

```

## 8.6.4 Packaging non-G2 Resources

Although it is recommended that Package Items refer to 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 the “raw” news objects would be as follows:

**LISTING 11 Simple Group Set using @href at CCL**

```

<groupSet root="G1">
  <group i d="G1" rol e="group: mai n">
    <i temRef href="http: //www. exampl e. com/NewsML-G2/tutori al -i tem-A. xml "
        contenttype="appl ication/ni tf+xml "
        format="fmt: ni tf"
        si ze="2345">
        <ti tle>Obama annonce son équi pe</ti tle>
    </i temRef>
    <i temRef href="http: //www. exampl e. com/NewsML-G2/tutori al -i tem-B. jpg"
        contenttype="i mage/j peg"
        format="fmt: j peg"
        si ze="300039">
        <ti tle>Barack Obama arri ve à Washi ngton</ti tle>
    </i temRef>
  </group>
</groupSet>

```

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

## 8.7 Packages and Links

Implementers may encounter the 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 reasons which are detailed below.

### 8.7.1 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 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.

### 8.7.2 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.

### 8.7.3 When to use <link>

The <link> property is used in the <itemMeta> section of a 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 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 G2 News Item conveys formatted text which references an illustration, a dependency link from the article to the illustration is indicated using <link>.

### 8.7.4 Link properties

Link uses the G2 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 extracted from the target resource may also be added.

At PCL the Link1Type datatype is used, optionally with more extensive attributes, which permits any property extracted from the target resource may be used as a processing hint. See [Hint and Extension Point](#).

#### 8.7.4.1 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="i rel: 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: "Derived From" and all other relationships in the CV.

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.

#### 8.7.4.1.1 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)

#### 8.7.4.2 Target Resource Attributes

See [6.3.3](#)

#### 8.7.4.3 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>
```

#### 8.7.4.4 Hint and Extension Point

At PCL, if the target resource is a G2 Item, any number of metadata properties of the resource 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 processing hints is subject to the following rules:

- ❖ In NewsML-G2 2.4, EventsML-G2 1.3, immediate child elements of <itemMeta> or <contentMeta> optionally with their descendants, may be used directly. For all other elements, the full path must be used, excluding the root element of the G2 Item.
- ❖ From NewsML-G2 2.5, EventsML-G2 1.4, 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 G2 Item.

The reason for these change is to avoid confusion when extracting G2 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](#) (below)



## 8.7.5 Link examples

### 8.7.5.1 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>
```

### 8.7.5.2 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 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="2.4"
  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>
  <contentSet>
    <inlineXML>
      <html xmlns="http://www.w3.org/1999/xhtml">
        ....
        <p>At Omaha Beach, President Obama led a ceremony to mark the landing of
          thousands of U.S. troops on D-Day.</p>
        
      </html>
    </inlineXML>
  </contentSet>
</newsItem>
```

### 8.7.5.3 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>

## 9 Concepts and Concept Items

### 9.1 Introduction

Concepts in 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 G2 Standards.

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.

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.

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

### 9.2 What is a Concept?

A Concept in G2 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" G2 concept expressions thus conform with an RDF triple of subject, predicate and object

Concepts may be either uncontrolled or controlled. Controlled concepts are managed by an "authority" (an organisation or company) and are maintained in Controlled Vocabularies. 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.

Every concept, whether controlled or uncontrolled must be identified, and the identifier used must be unique in its scope.

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

This Chapter will describe Controlled Concepts.

### 9.3 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 15](#) it uses the same G2 template as News Item ([Figure 4](#)) and Package Item ([Figure 7](#)) 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 is should be machine-readable.

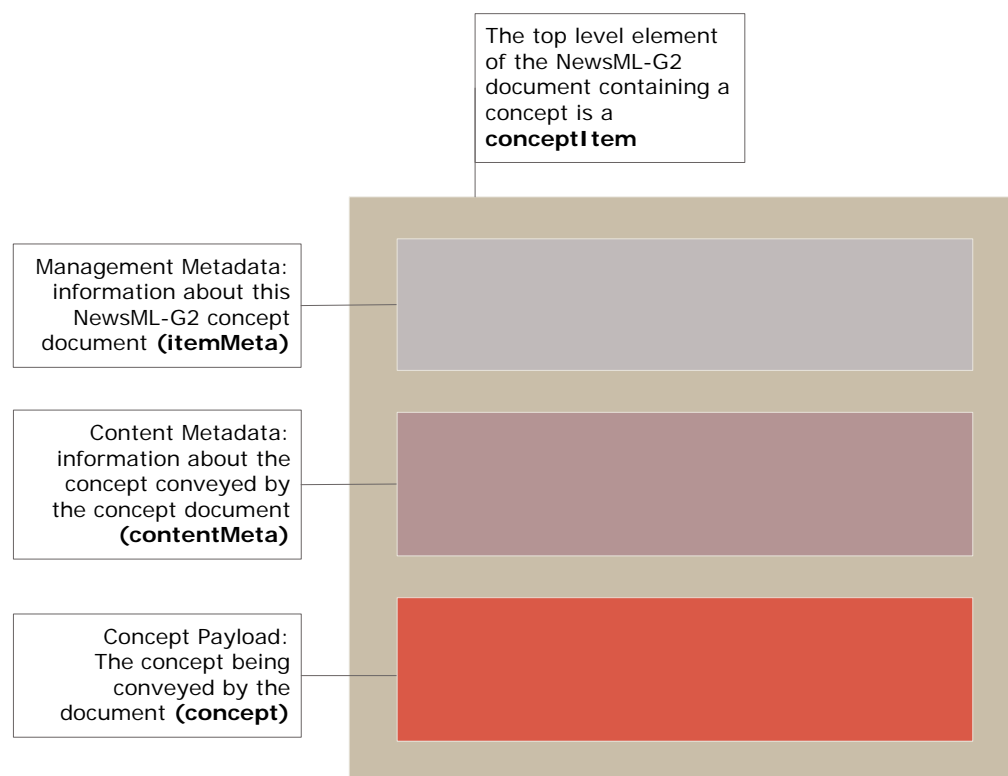


Figure 15: Concept Items share the same template as News Items and Package Items

## 9.4 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:

### 9.4.1 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 continue to be able to resolve the concept.

### 9.4.2 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>
```

### 9.4.3 Concept Type <type> and Facet <facet>

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

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 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).

A <facet> uses either a @qcode or @literal to additionally describe other inherent characteristics of a concept in terms of a named relationship with another concept. If the related concept is identified by a QCode; each provider should provide its own CV for this purpose (discussed in [Knowledge Items](#))

The relationship may be identified in the @rel attribute by a QCode; in this case a controlled vocabulary (CV) of relationships would also be required.

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

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

And “Jean-Claude Trichet” (Subject) “has an occupation of” (Predicate) “public official” (Object):

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

### 9.4.4 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 human-readable 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'études 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.

### 9.4.5 Note <note>

Additional human-readable 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>

```

### 9.4.6 Remote Info <remoteInfo>

This is a Link Type (CCL) or Link 1 Type (PCL) wrapper has the same semantic as the <link> property (see 5.4.5 and 7.2.1.9), 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](#)

## 9.5 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 G2 document may have a concept for Germany identified by the provider's QCode "country:de". Some recipients may have standardized on using ISO-3166 Country Codes to classify nationality. The provider can assist recipients to make a direct reference to their preferred scheme using "sameAs":

```

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

```

<related> allows the expression of a relationship with 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>
```

## 9.6 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.

### 9.6.1 Person Details <personDetails>

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

#### 9.6.1.1 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 values may be truncated, starting on the right (seconds) according to the precision required, but MUST, at minimum, have a year, for example:

```
<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.

#### 9.6.1.2 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.





### 9.6.1.3 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>&lt;email role="addressrole: office"&gt;   info@ecb.eu &lt;/email&gt;</pre>
Instant Message Address	<im>	Electronic Address	<pre>&lt;im role="imsrvc: reuters"&gt;   j.c.trichet.ecb.eu@reuters.net &lt;/im&gt;</pre>
Phone Number	<phone>	Electronic Address	
Fax Number	<fax>	Electronic Address	
Web site	<web>	IRI	<pre>&lt;web role="webrole: corporate"&gt;   www.ecb.eu &lt;/web&gt;</pre>
Postal Address	<address>	Address	See below for details of Address properties. The Address may have a @role to denote the type of address it 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”

### 9.6.1.4 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="addressrole: 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>
```



## 9.6.2 Organisation Details <organisationDetails>

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

### 9.6.2.1 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 [9.6.1.1](#))

### 9.6.2.2 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="Ictypes: headoff" literal="Frankfurt"/>
```

Or

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

### 9.6.2.3 Contact Information <contactInfo>

Contact information associated with the organisation, uses the same structure as described in [9.6.1.4](#).

## 9.6.3 Geopolitical Area Details

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

### 9.6.3.1 Position <position>

This expresses the coordinates of the concept using the following attributes

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

## 9.6.4 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 is has different properties to a purely-geographical point. POI may have the following additional properties:



#### 9.6.4.1 Address

The location of the point of interest expressed as a postal address. The <address> element is a wrapper for child elements described in 9.6.1.4. 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.

#### 9.6.4.2 Position <position>

The coordinates of the location as described in 9.6.3.1

#### 9.6.4.3 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>
```

#### 9.6.4.4 Capacity <capacity>

The capacity of the POI is expressed as a Label:

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

#### 9.6.4.5 Contact Information <contactInfo>

Contact information for the POI uses the <contactInfo> structure as described in 9.6.1.3.

#### 9.6.4.6 Access details <access>

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

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

#### 9.6.4.7 Detailed Information <details>

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

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

### 9.6.5 **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 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:

#### 9.6.5.1 Creation Date <created>

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

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



### 9.6.5.2 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.

### 9.6.5.3 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>
```

## 9.7 Code listings

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

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

### LISTING 12 "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="081127185821"
  standard="NewsML-G2"
  standardversion="2.2"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    NAR_1.3-spec-ConceptItem-Core.xsd">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <rightsInfo>
    <copyrightHolder>
      <name>IPTC - International Press Telecommunications Council, 20
        Garri ck Street, London WC2E 9BT, UK</name>
    </copyrightHolder>
    <copyrightNotice>Copyright 2008, IPTC, www.iptc.org, All
      Rights Reserved</copyrightNotice>
  </rightsInfo>
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
    <versionCreated>2008-11-27T18:58:21+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>
```



```
<defini ti on xml : l ang=" / i t">Creazi one e rappresen tazi one del l'opera d'arte, gli
Interessi intellettuali, il gusto e le emozioni umane</defini ti on>
<note xml : l ang=" / i t">nessuno</note>
</concept>
</conceptI tem>
```

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 G2 Items containing references to the Concept can be resolved:

### LISTING 13 "Named Entity" Concept

```
<?xml versi on="1.0" encodi ng="UTF-8"?>
<conceptI tem xml ns="http://i ptc.org/std/nar/2006-10-01/"
xml ns:xsi="http://www.w3.org/2001/XMLSchema-i nstance"
gui d="urn:newsml:i ptc.org:20080229:ncdci -person"
versi on="081127185821"
standard="NewsML-G2"
standardversi on="2.2"
xsi:schemaLocati on="http://i ptc.org/std/nar/2006-10-01/
NAR_1.3-spec-ConceptI tem-Core.xsd">
<catalogRef
href="http://www.i ptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
<ri ghtsI nfo>
<copyri ghtHol der>
<name>IPTC - International Press Telecommuni cations Council, 20
Garri ck Street, London WC2E 9BT, UK</name>
</copyri ghtHol der>
<copyri ghtNoti ce>Copyri ght 2008, IPTC, www.i ptc.org, All
Ri ghts Reserved</copyri ghtNoti ce>
</ri ghtsI nfo>
<i temMeta>
<i temCl ass qcode="ci nat:concept" />
<provi der l i teral="IPTC" />
<versi onCreated>2008-12-27T14:48:25Z</versi onCreated>
<fi rstCreated>2008-12-29T11:00:00Z</fi rstCreated>
<pubStatus qcode="stat:usable" />
<ti tle xml:l ang="en">Concept I tem descri bi ng Jean-Claude Tri chet</ti tle>
</i temMeta>
<concept>
<conceptId created="2009-01-10T12:00:00Z" qcode="peopl e:329465" />
<type qcode="cpnat:person" />
<name xml:l ang="en-GB">Jean-Claude Tri chet</name>
<defini ti on xml:l ang="en-GB" rol e="drol:bi og">
Jean-Claude Tri chet (born 20 December 1942) is a French civil servant who is
the current president of the European Central Bank since 2003.<br />
Tri chet was born in Lyon, France and educated at the École des Mines de Nancy.
He later trained at the Institut d'etudes poli tiques de Paris and the École
nati onale d'admini stration, two higher educati on i nsti tuti ons wi th a
tradi ti on of provi di ng France's poli ti cal and state admini strati ve
el i te.<br />
In 1993 Tri chet 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 />
</defini ti on>
<note xml:l ang="en-GB" rol e="nrol:di sambi guati on">
Not Jean-Claude Tri chet, the i nternati onal powerboat racer
</note>
<facet rel="relati on:occupati on" qcode="jobtypes:puboff" />
<sameAs type="otherprov:AFP" qcode="pers:567223">
<name>TRI CHET, Jean-Claude</name>
</sameAs>
<personDetail s>
<born>1942-12-14</born>
<affiliati on type="orgrol:empl oyer" qcode="org:ECB">
<name>European Central Bank</name>
</affiliati on>
<contactI nfo rol e="contactrol e:offi ci al">
<email rol e="addressrol e:offi ce">i nfo@ecb.eu</email>
<i m rol e="i msrvc:reuters">j.c.trichet.ecb.eu@reuters.net</i m>
<phone rol e="phonerol e:swi tch">+49 69 13 44 0</phone>
<fax rol e="faxrol e:central">+49 69 13 44 60 00</fax>
<web>www.ecb.eu</web>
<address rol e="addrol e:physi cal">
<li ne>Kai serstrasse 29</li ne>
<locali ty l i teral="Frankfurt am Mai n" />
<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>

```

## 9.8 Supplementary information about a Concept

Links can be used to enhance the information carried by a 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, NewsML-G2 version 2.3 and EventsML-G2 1.2 introduced a <remoteInfo> property to <concept>, with a datatype of LinkType (CCL) and Link1Type (PCL), matching that of <link>. This enables providers to carry 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="2.4">
  <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>
    <facet rel="relation:occupation" qcode="jobtypes:puboff" />
    <remoteInfo
      rel="irel:seealso"
      contentType="image/jpeg"
      resDref="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
      <title>
        ECB official portrait picture of Jean-Claude Trichet
      </title>
    </remoteInfo>
    <personDetails>
      <born>1942-12-14</born>
      .....
      <contactInfo role="contactrole:official">
        .....
      </contactInfo>
    </personDetails>
  </concept>
  .....
</conceptItem | knowledgeItem>

```

← NO <link>

← “link” start

← Item ref

← “link” end



Using the rules given in **Hint and Extension Point** when copying a property from the target G2 Item, the parent property must be included if it is not either <contentMeta> or <itemMeta>:

```
<remoteInfo
  rel="related:seeAlso"
  contentType="video/mpeg">
  residence="tag:acmenews.com,2008:TX-PAR:20090529:JYC90"
<description>
  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>
</partMeta>
<partMeta partId="part2" seq="2">
  <description>The second part shows...</description>
</partMeta>
</remoteInfo>
```

Annotations in the original image:

- An arrow points from the `residence` attribute to the text "no parent".
- An arrow points from the `<description>` element to the text "includes parent".

## 9.9 Concepts and Controlled Values

### 9.9.1 Concept Identifier and Concept URI

The concept conveyed in **LISTING 13** 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 G2 Concept that is represented by a QCode<sup>8</sup> 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: the G2 term is a "Concept-URI". G2 specifies that it must take the URI subtype of 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
```

If this URL is entered into a Web browser, the page shown in **Figure 16** is returned:

<sup>8</sup> Concepts which are local to the scope of a G2 document may be represented by a literal value.



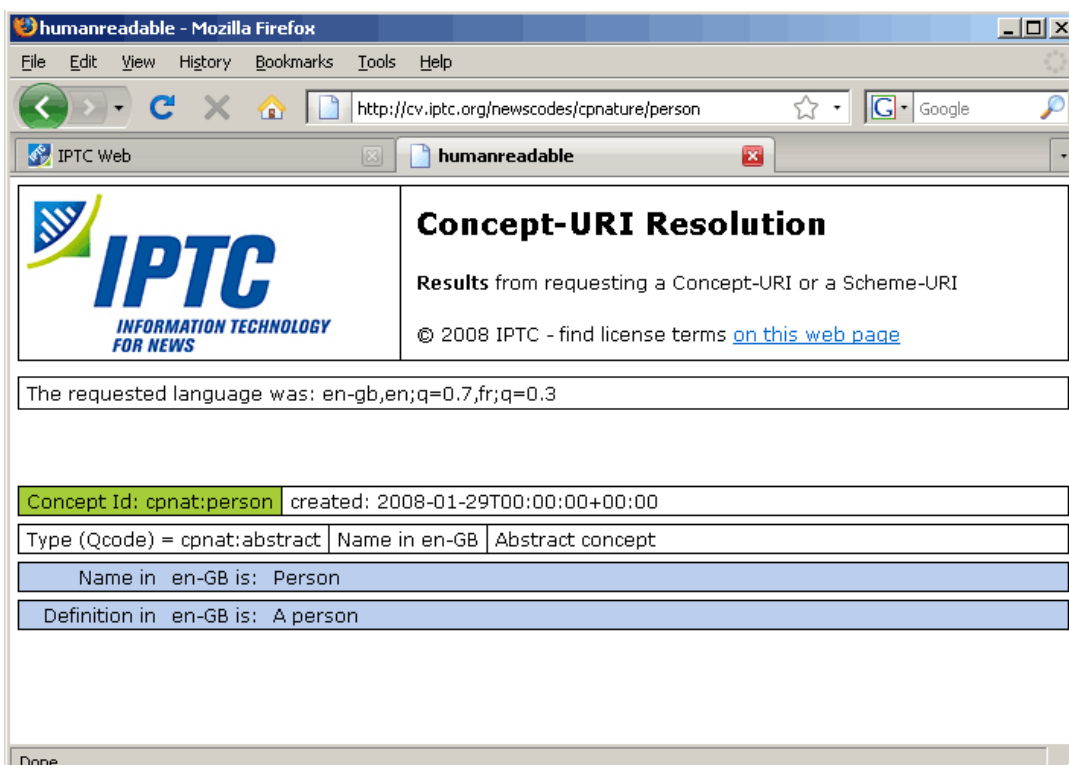


Figure 16: Entering the Concept URI in a browser returns human-readable information

## 9.9.2 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.

## 9.9.3 Resolution of QCodes

Scheme URIs are contained in <catalog> statements in the root element of a G2 document. A G2 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 NewsCodes which are required by the G2 Standards 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 local G2 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/pubstatusg2/" />
</catalog>
```

Use the Catalog information, a G2 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.

## 9.10 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 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](#).

## 10 Knowledge Items

### 10.1 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.

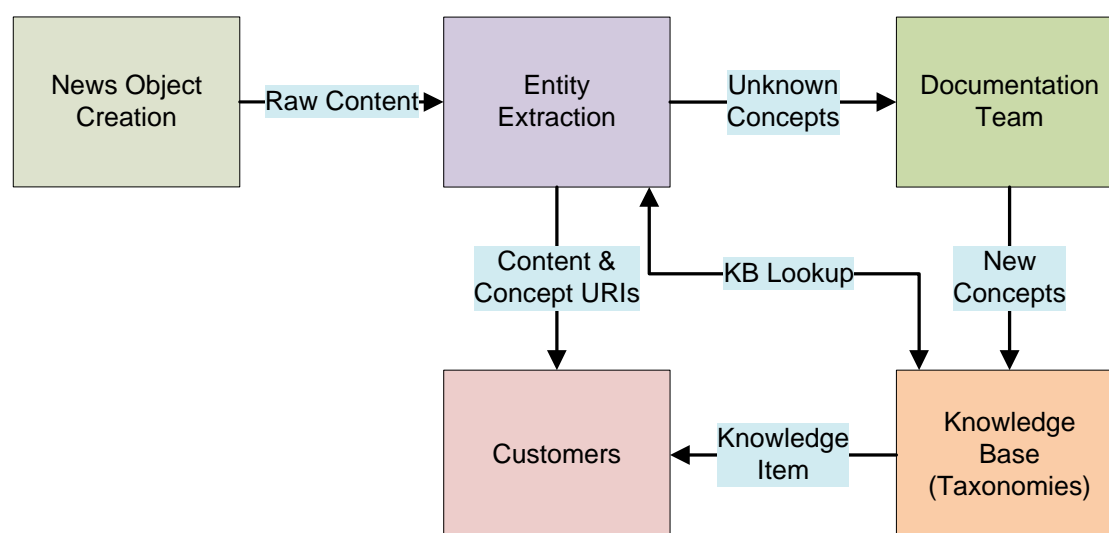
There will also be references to abstract notions such as subject classifications that enable the news event to be searched and sorted according to user preferences.

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**, 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 these personalities that can be updated and referred to over time.

**Figure 17** shows a possible information flow for news information that exploits these possibilities.



**Figure 17: 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. In G2, 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" />
  <type qcode="cpnat: abstract" />
  <name xml:lang="en-GB">
    Text Item(s)
  </name>
  <definition xml:lang="en-GB">
    Text content in a News/PackageItem
  </definition>
</concept>
```

matching QCode

Name

Definition

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.

## 10.2 Structure and Properties

Knowledge Items follow the NAR anyItem model and so share a common structure with News Items, Package Items and Concept Items. This is shown diagrammatically in [Figure 18](#) on the following page.

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

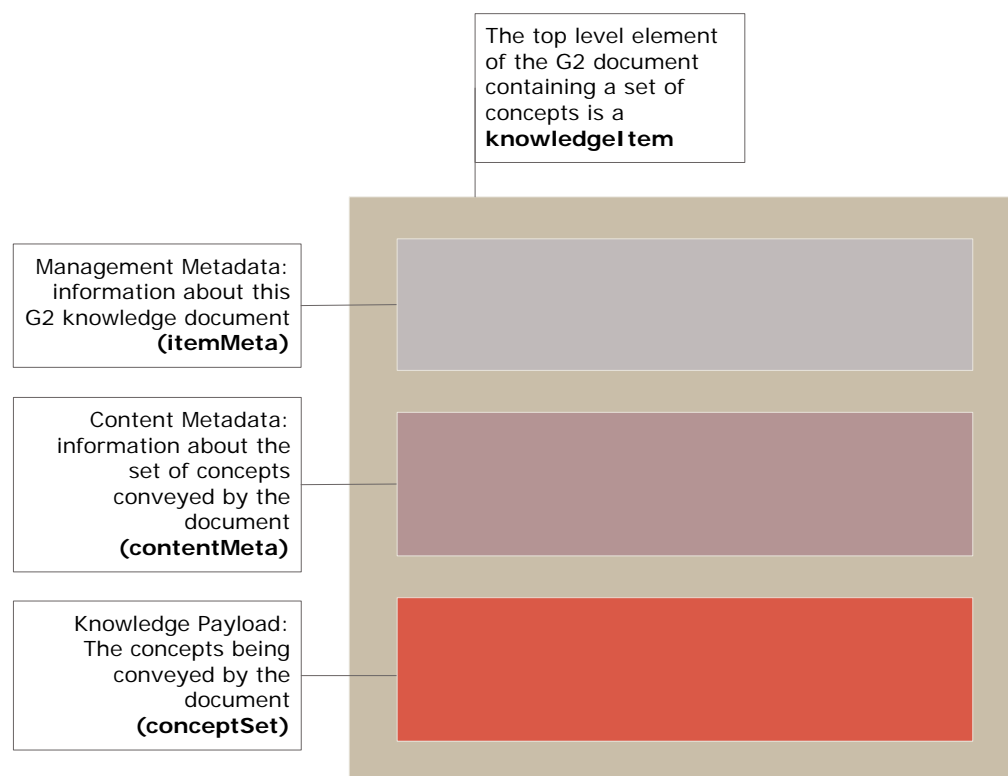


Figure 18: Structure of a Knowledge Item mirrors that of other G2 Items

### 10.2.1 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/
    NAR_1.3-spec-KnowledgeItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090126:qqwpi ruuew4714"
  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_6.xml" />
```

XML Schema  
GUID  
and version

Remote Catalog  
reference

### 10.2.2 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](#).



### 10.2.3 Content Metadata

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

#### 10.2.3.1 Administrative Metadata

Some date-time information may be needed:.

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

#### 10.2.3.2 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>
```

### 10.2.4 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 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 G2 infrastructure.

### 10.3 Use Case: a Controlled Vocabulary for <accessStatus>

One of the available G2 properties for describing a news event is <accessStatus> which is used in **EventsML-G2** 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 G2 Specification recommends that when a receiving application processes the QCode found in a G2 document, it **SHOULD** be able to resolve it to a resource (or resources) containing information about the concept that the QCode represents.<sup>9</sup>

If a customer receives event information containing the Access Status property and a QCode of "access:difficult", it is highly 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 way<sup>10</sup> 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 could be made available as a web-accessible 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

<sup>9</sup> Processing rules and recommendations are detail in Chapter 9 of the G2 Specification Document: Dealing with Controlled Values, available for download from the IPTC web site [www.iptc.org](http://www.iptc.org)

<sup>10</sup> Not the only method – 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	Access is difficult	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
	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 14 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/
    NAR_1.3-spec-KnowledgeItem-Core.xsd"
  guid="urn:newsml:iptc.org:20090202:ncdki-accesscode" version="20090202"
  standard="NewsML-G2"
  standardversion="2.2" >
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
```

```

<versionCreated>2009-01-28T16:00:00Z</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 a peut-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
    </definition>
  </concept>

```



behindern große oder schwere Gegenstände. Beraten Sie mit dem Veranstalter in  
Verbindung  
</definition>  
</concept>  
</conceptSet>  
</knowledgeItem>



# 11 Controlled Vocabularies and QCodes

## 11.1 Introduction

One of the fundamental ideas underpinning the G2 Standards 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 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 the G2-Standards, 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 G2 Item to form a URI that is unique within the scope of the Item.

## 11.2 Business Case

Controlled Vocabularies are needed in information exchange because they establish a common ground for understanding content that is language-independent. G2 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.

## 11.3 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<sub>dec</sub> = 003A<sub>hex</sub>)
- ❖ 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 G2 documents, but with an important distinction in their application, as follows:

- ❖ Many G2 property attributes require a concept, represented by a QCode. For example the values of @role and @type attributes are represented by a QCode. These REFINES the value of the property.
- ❖ However, when a G2 property has a @qcode attribute, 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).

### 11.3.1 Scheme Alias to Scheme URI

The key to resolving scheme aliases is the Catalog information, a child of the root element of every 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 G2 processor now has a **Scheme URI** that can be used as the next step in resolving the QCode.

#### 11.3.1.1 Catalogs

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

- ❖ directly in a 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 a G2 item:

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

### 11.3.2 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 G2 Specification documents describe how an http-URL should be resolved.

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

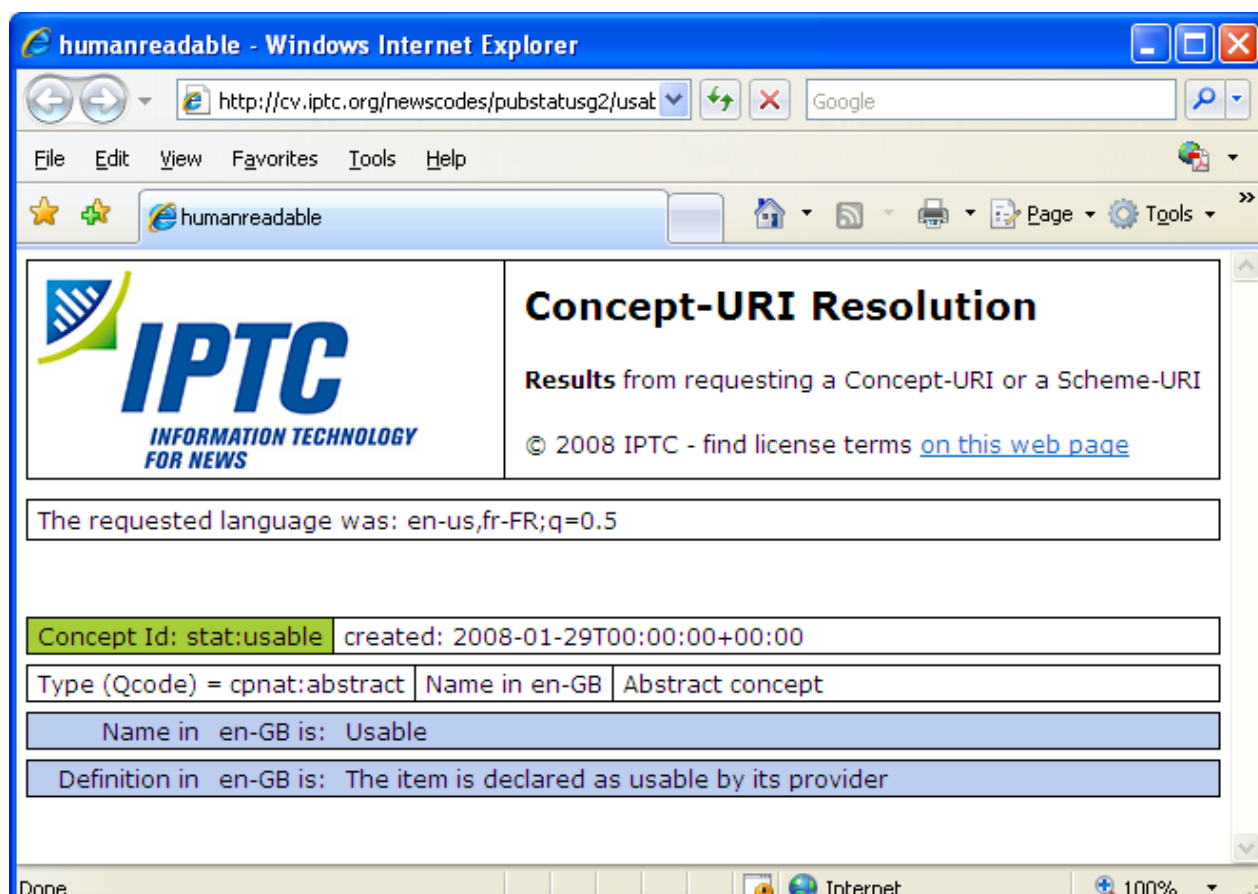


Figure 19: Human-readable browser page of a Concept URI

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 G2 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="topic: 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 G2 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 G2 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.

## 11.4 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 G2 processor might encounter the following XML in a 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, VI ad i mi r VI ad i mi rovi tch
Name (known as)	VI ad i mi r VI ad i mi rovi tch Put i 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.





## 11.5 Managing Controlled Vocabularies as G2 Schemes

### 11.5.1 Knowledge Items

In a workflow where partners are exchanging news information using G2 Standards, Knowledge Items are the most G2-compliant method of distributing Controlled Vocabularies. The sections below describe the steps to create a new CV: first by creating a Scheme (11.5.2) and next creating a Knowledge Item from a set of existing Concept Items (11.5.3) 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 G2 Knowledge Item. It is sufficient, for the correct processing of a 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 in a G2 environment. Other means such as paper, fax or email are permissible but at a price of less efficient automated G2 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.

### 11.5.2 Creating a new Scheme

A G2 controlled vocabulary is a set of concepts. To create a CV as a 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 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"`

### 11.5.3 Creating a new Knowledge Item for distributing a CV

A G2 Knowledge Item contains concepts from one to many Schemes. To create a KI:

- ❖ Identify the set of G2 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 10](#) for more information on the structure of Knowledge Items.

## 11.5.4 Maintaining a CV

### 11.5.4.1 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 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 G2 Items that pre-date the changes. Use of @retired ensures that G2 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

### 11.5.4.2 Changes to Catalogs

Catalog files only need to be changed when new Controlled Vocabularies are created and included in G2 Items. Changes to the Scheme contents do not need to be reflected in a changed Catalog file, or <catalog> element within the G2 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 at the <http://www.iptc.org/goto?G2cataloginfo> which also provides a link to download the latest file.



Receiving applications MUST use the catalog information contained in the G2 file 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.

## 11.6 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 G2 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.

G2-Standards require that remote catalogs – the file(s) that map Scheme Aliases to Scheme URIs – are retrieved by G2 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 G2 Items that pre-date the catalog changes can continue to be processed using the previous catalog URI.

### 11.6.1 Resolving Scheme Aliases

Some G2 properties are important for the correct processing of a G2 Item, for example the Item Class property tells a receiving application the type of content being conveyed by a G2 News Item: text, picture, audio etc, using the News Item Nature NewsCodes.. (Concept Items use the complementary Concept Item Nature CV). A G2 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. But 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 G2 model: consistency of scheme aliases between different providers – or even by the same provider – cannot be guaranteed, and in G2 they do not have to be guaranteed. For this reason it would be unwise for implementers of G2 processors 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 a G2 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 G2 content is routed to the appropriate department, according to the Subject NewsCodes found in the G2 Items,
- ❖ A routing table is set up in the G2 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 G2 processing.

In the example, the catalog used by the G2 Item resolves the scheme alias "sc" contained in the QCode. The G2 Item contains the line pointing to the catalog file:

```
<catalogRef href="http://www.example.org/std/catalog/catalog.example_10.xml" />
```

The G2 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 G2 Item processed appropriately.

### 11.6.2 Resolving Concept URIs

The IPTC recommends that Schemes SHOULD resolve to a Web resource, and that Scheme Authorities who disseminate news using the G2-Standards should make their Schemes available as Knowledge Items.



#### 11.6.2.1 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.

#### 11.6.2.2 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 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.

#### 11.6.2.3 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. As noted above, a Scheme Authority may not guarantee the availability and capacity of connections to its hosted Knowledge Items.

From the receiver's point of view, it may be unwise to have business-critical news applications that rely on a third-party system beyond the receiver's control.

G2 processors are therefore recommended to retrieve and cache the contents of available Knowledge Items. The cache may be regularly updated, and the providers should advise their customers on the recommended frequency of updates.

This is not a new issue: most news providers have Controlled Vocabularies that pre-date G2, for example in IPTC 7901, and there are channels and conventions for advising customers of changes to CVs. 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. News organisations can subscribe to an RSS feed that notifies of changes to the IPTC Schemes.

### 11.7 Private versions of CVs

In NewsML-G2 2.5 and EventsML-G2 version 1.4 providers can create their own versions of other CVs via a "sameAs" relationship. This solves some issues that some news providers have encountered, such as adding translations of free-text properties (name, definition, note etc), which are not available with the original scheme, or adding additional information, e.g. usage notes.

#### 11.7.1 **Example Business Case**

- ❖ 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.
- ❖ The national news agency of Country X would like to provide its customers with a name and definition of the IPTC Subject NewsCodes in their local language.

#### 11.7.2 **Solution**

An optional child property <sameAs> will be added to the <scheme> element in the <catalog> wrapper. The <sameAs> will be of IRIType and will be repeatable. This will enable providers to create their own CV,



and assert that the members of this Scheme are the “same as” the corresponding members of another Scheme or Schemes, specifying the Scheme URIs.

The semantics of <sameAs> are “all of the concepts in the scheme identified by @uri have a ‘same as’ relationship to concepts with the same code in the original scheme identified by this <sameAs>”. Thus:

```
<catalog xmlns="http://iptc.org/std/nar/2006-10-01/"
  additionalInfo="http://www.iptc.org/goto?G2cataloginfo">
  <scheme alias="acdc" uri="http://cv.example.org/ourcodes/audio/"> ← private CV
    <sameAs>http://cv.iptc.org/newscodes/audiocodec/</sameAs> ← original CV
    <sameAs>http://catalog.ebu.ch/audio/</sameAs> ← original CV
  </scheme>
</catalog>
```

- ❖ The Scheme identified by @uri (the **provider's** private scheme) must NOT use a code that does not exist in ALL of the **original** Schemes identified by the <sameAs> elements
- ❖ Some codes and concepts in the original Schemes MAY not exist in the provider's Scheme, for example if the original Scheme has new terms added which are not yet in the provider's Scheme..
- ❖ Therefore one cannot add new concepts to the provider's Scheme that have the effect of extending the original Schemes.
- ❖ A concept identified by a code in the provider's Scheme MUST be semantically equivalent to the corresponding concept, identified by the same code, in the original Schemes.

The Scheme <sameAs> complements the <sameAs> child element of a Concept. It enables a provider to apply a <sameAs> relationship at the level of a set of Concepts as well as at the level of individual Concepts.

This speeds up QCode processing: a G2 processor does not have to check the individual concept for its sameAs relationships but can apply this relationship directly if the scheme identifier of the concept (used as property value) matches the scheme identifier with the sameAs child.

## 11.8 Complementary CVs

An organisation may identify a business need for additional concepts to be added to a CV maintained by another Scheme Authority. For example, a receiver may be using another organisation's CV for identifying the stages in a shared workflow. The receiver would have two courses of action:

- a) Apply to the Scheme Authority for new concepts to be added to the CV. For example, IPTC members are entitled to request the addition and/or retirement of terms in IPTC Schemes with the agreement of other members.
- b) Create a new scheme that complements the original scheme, but uses properties such as <broadener> and <sameAs> to link the new scheme to the original scheme. A concept is the <sameAs> another concept if its semantics are the same but it contains more details, such as a translation in another language. A concept with a <broadener> relationship to another concept is a new concept with semantics narrower than those of the <broadener> concept.

Example: Original Scheme A

```
<conceptSet>
  <concept>
    <conceptId qcode="A: C1" />
    ...
  </concept>
  <concept>
    <conceptId qcode="A: C2" />
    ...
  </concept>
</conceptSet>
```

Example: Complementary Scheme B

```
<conceptSet>
  <concept>
    <conceptId qcode="B: C1" />
    <sameAs qcode="A: C1" />
    ...
  </concept>
  <concept>
    <conceptId qcode="B: C2" />
    ...
  </concept>
</conceptSet>
```





```
<sameAs qcode="A: C2" />
...
</concept>
<concept>
  <conceptId qcode="B: C3" />
  <broader qcode="A: C2" />
  ...
</concept>
</conceptSet>
```

## 11.9 QCodes and non-URI characters

The G2 Standards specify 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 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 if resolving the code to a Concept URI

### 11.9.1 Non-ASCII characters

The encoding described above assumes that the character(s) to be per-cent encoded are from the US-ASCII character set (consisting of 94 printable characters plus the space).

If a code contains non-ASCII characters, for example accented characters, the Unicode encoding UTF-8 must be used, in line with normal practice.

For example, the UTF-8 encoding of the “å” character is a two-byte value of C3<sub>hex</sub> A5<sub>hex</sub> which would be percent-coded as “%C3%A5”.



## 11.9.2 White Space in Codes

Many codes in existing controlled vocabularies contain spaces. As the G2 Specification does not allow white space characters in G2 Codes, this section recommends a workaround.

Whitespace characters in Codes - in practice, only spaces (20<sub>hex</sub>) - 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.

### 11.9.2.1 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 G2 using "~" to replace spaces: "ab~3x"

Receiver decodes G2 code for matching against the existing CV: "ab 3x"

## 11.10 Syntactic Processing of QCodes

This section provides a summary of the processing model. Please also read Chapter 9 of the G2 Specification (download from [www.iptc.org](http://www.iptc.org)) for a full technical description.

### 11.10.1 Creating QCodes from Scheme Aliases and Codes

#### 11.10.1.1 Scheme Aliases

These do not have to be encoded as they will never be part of the full Concept URI. A Scheme Alias may contain any character except a colon (3A<sub>hex</sub>) or white space characters (20<sub>hex</sub> or 09<sub>hex</sub> or 0D<sub>hex</sub> or 0A<sub>hex</sub>)

#### 11.10.1.2 Codes

In NewsML-G2 2.4 and NewsML-G2 1.3, the IPTC recognised the problem of legacy codes containing characters that require encoding before use within a URI, and made the processing model for Schemes and QCodes more explicit:

To create a QCode for a G2 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" />

### 11.10.2 Processing Received Codes

To resolve a QCode received in a G2 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 G2 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: bâr → 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.

## 11.11 QCode and Literal Identifiers

The G2-Standards recognise that it is not always possible to use Controlled Values, i.e. QCodes, as identifiers. For example, it may be impractical to store fine-grained metadata about locations such as city neighbourhoods in a CV, although the cities themselves may be part of a scheme.

For this reason, the G2 designers conceived Flexible Property datatypes, that may have EITHER a @qcode or a @literal value.

### 11.11.1 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 G2 Items.
- ❖ A @literal is an unambiguous identifier only within the scope of the containing G2 Item, but is not a globally unique identifier and CANNOT be shared with other G2 Items.

### 11.11.2 What @literal is; what it is NOT

A @literal is an identifier, which may optionally be intelligible to the human reader but is intended to be processed by software; it is not intended to be a human-readable label. If a concept identified by @literal is intended for display, providers SHOULD add the human-readable <name> property. For example:

```
<subject type="cpnat: poi " I i t e r a l ="e i f f e l t o w e r">
  <name xml:lang="en">The E i f f e l T o w e r</name>
  <name xml:lang="fr">La T o u r E i f f e l</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.

### 11.11.3 Use of @literal in a G2 Item

A @literal value unambiguously identifies a concept within the scope of a G2 Item. The following is a valid use of @literal:

```
<contentMeta>
  . . .
  <located type="cpnat: poi " I i t e r a l ="i n t 0 0 1" />
  . . .
  <subject type="cpnat: poi " I i t e r a l ="i n t 0 0 1" />
  . . .
</contentMeta>
<assert I i t e r a l ="i n t 0 0 1">
  <name xml:lang="en">The E i f f e l T o w e r</name>
  <name xml:lang="fr">La T o u r E i f f e l</name>
</assert>
```

The @literal properties of both <subject> and <located> refer to the same inline concept (<assert>) identified by a @literal value "int001".

The following example would be INCORRECT:

```
<contentMeta>
  . . .
  <located type="cpnat: geoArea" I i t e r a l ="i n t 0 0 1">
    <name>P a r i s</name>
  </located>
  . . .
  <subject type="cpnat: poi " I i t e r a l ="i n t 0 0 1">
```



```
<name>The Ei ffel  Tower</name>
</subj ect>
</contentMeta>
```

The @literal value “int001” is being used to identify two different concepts within the same G2 Item.



This page intentionally blank



## 12 EventsML-G2

### 12.1 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 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 to 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: the need to express the basic news agenda of “what, when, where and who” combined with notions of how news organisations respond to news events, such as job assignments and content planning.

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
- ❖ planned or actual journalistic tasks and products for a news event – the event “coverage”.

#### 12.1.1 Business Advantages of adopting EventsML-G2

EventsML-G2 is more than a planning tool. Because it shares the G2 architecture, an event managed in a G2 framework 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 G2 QCode. When content related to this event is created (text, pictures, audio and video), G2 enables 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.

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 iCalendar<sup>11</sup> specification that is

---

<sup>11</sup> A mapping of iCalendar properties to EventsML-G2 properties can be found in the G2 Specification document on the IPTC web site ([www.iptc.org](http://www.iptc.org)) See also [IETF iCalendar Specification](http://www.iptc.org)

supported by popular calendar and scheduling applications such as Microsoft Outlook, Lotus Notes, 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 may be inherited by the news content from the 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.

### 12.1.2 Structure and compatibility

EventsML-G2 is part of the G2 family derived from the News Architecture (NAR) data model. Thus it inherits its structure and components from the NAR “anyItem” class, as does NewsML-G2 and SportsML-G2, and is extended only to provide a few specific event-related features.

- ❖ Events use the same identification and versioning properties as other G2 objects
- ❖ The <itemMeta> block holds management information about the G2 object conveying the event information.
- ❖ The <contentMeta> block holds common administrative metadata about the event, or events, conveyed by the G2 object.

There are two methods for conveying events in G2, each suited to a particular type of information, or application:

- ❖ Volatile “standalone” event information may be conveyed using NewsML-G2 using an <events> structure inside a News Item. ([Figure 20](#))
- ❖ Persistent event information that may be referenced by other events and by other G2 objects is conveyed in a Concept Item ([Figure 21](#)), or as part of a collection of events in a Knowledge Item.

The implementations are extremely similar – most of the properties of an event are shared by the two models – with the key difference that an event expressed as a Concept has a Concept ID, which makes it persistently and unambiguously identifiable.

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.

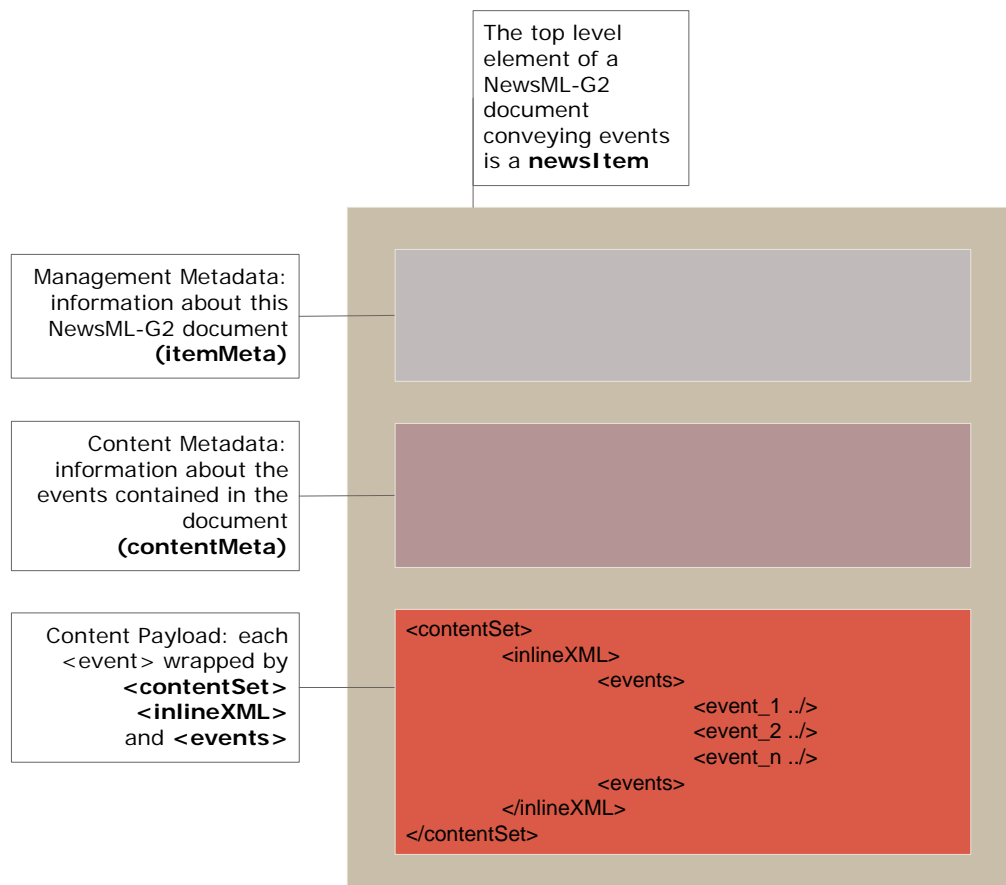


Figure 20: Multiple event information carried as `<events>` in a News Item

The use-case for managed 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 identifier, and this enables linking and navigation of content related to news stories.

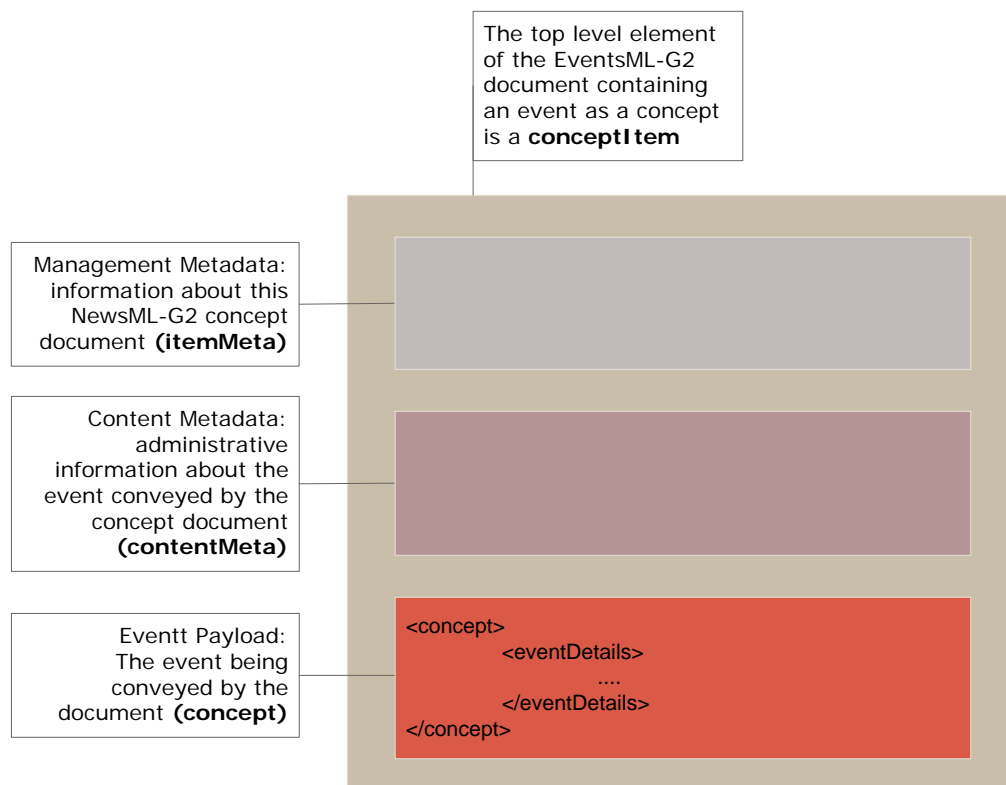


Figure 21: An event conveyed as a Concept using `<eventDetails>`



## 12.2 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, EventsML-G2 has a flexible set of properties that can convey these details. There is also a fully-featured date-time structure to express event occurrences, which conforms to the iCalendar specification.

### 12.2.1 What is the Event?

In order to convey event information, we first need to describe “what” the event is. In EventsML-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, Facets (some inherent 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.

### 12.2.2 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.

### 12.2.3 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.

### 12.2.4 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 Organisation and Contact Information properties can give precise details of the people and organisations responsible for an event, and how to contact them.

## 12.3 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 and considered important, which need detailed planning and direction of resources.
- ❖ Unplanned news – events that come to the attention of news organisations during the course of daily operations, for example triggered 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 News Coverage property may be used to inform customers what content they may expect to receive and if necessary the disposition of staff and resources. At CCL this is simply a set of repeatable Block type <edNote> elements that provide a natural language description of the coverage.

At PCL, it is possible to provide rich information structures describing the Content Type(s) to be delivered, with associated Descriptive Metadata, job assignments, the scheduled date-time of delivery, and the service on which the content will be provided.

## 12.4 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.

### 12.4.1 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, one with a @role of “short” and the other “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 Facet, the inherent characteristics of the Event can be expressed using QCodes to reference a controlled vocabulary (CV). One CV may list the types of event, such as “meeting”, “parliamentary session”, “music concert” and have values to indicate whether the event is “open to the public” or “private”.

Property	Type	Notes/Example
<name>	Intl String	<pre>&lt;name xml:lang="en"&gt;   Bank of Engl and Monetary Policy Commi ttee &lt;/name&gt;</pre>
<definition>	Block	<pre>&lt;defini ti on xml:lang="en" role="drole: short"&gt;   Monthly meeting of the Bank of Engl and commi ttee that   dec ides on bank lending rates for the UK. &lt;/defini ti on&gt; &lt;defini ti on xml:lang="en" role="drole: long"&gt;   The Bank of Engl and Monetary Policy Commi ttee meets each   month to decide on the minimum rate of interest that will   be charged for inter-bank lending in the UK financial   markets. &lt;br /&gt;   The commi ttee 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. &lt;/defini ti on&gt;</pre>
<facet>	Typed QCode	<pre>&lt;facet qcode="efacet: meet" /&gt; &lt;facet qcode="efacet: pri v" /&gt;</pre> <p>The property is repeatable and may have a @rel attribute to indicate the relationship to the characteristic. If omitted, the default relation is “is a”.</p>

<note>	Block	<p>A repeatable natural language note of additional information about the event, with an optional @role:</p> <pre>&lt;note role="role:toeditors"&gt;   Note to editors: an embargoed press release of the   minutes of the meeting will be released by the COI   within two weeks. &lt;/note&gt; &lt;note role="role:general"&gt;   The meeting was delayed by two days due to illness. &lt;/note&gt;</pre>
--------	-------	---

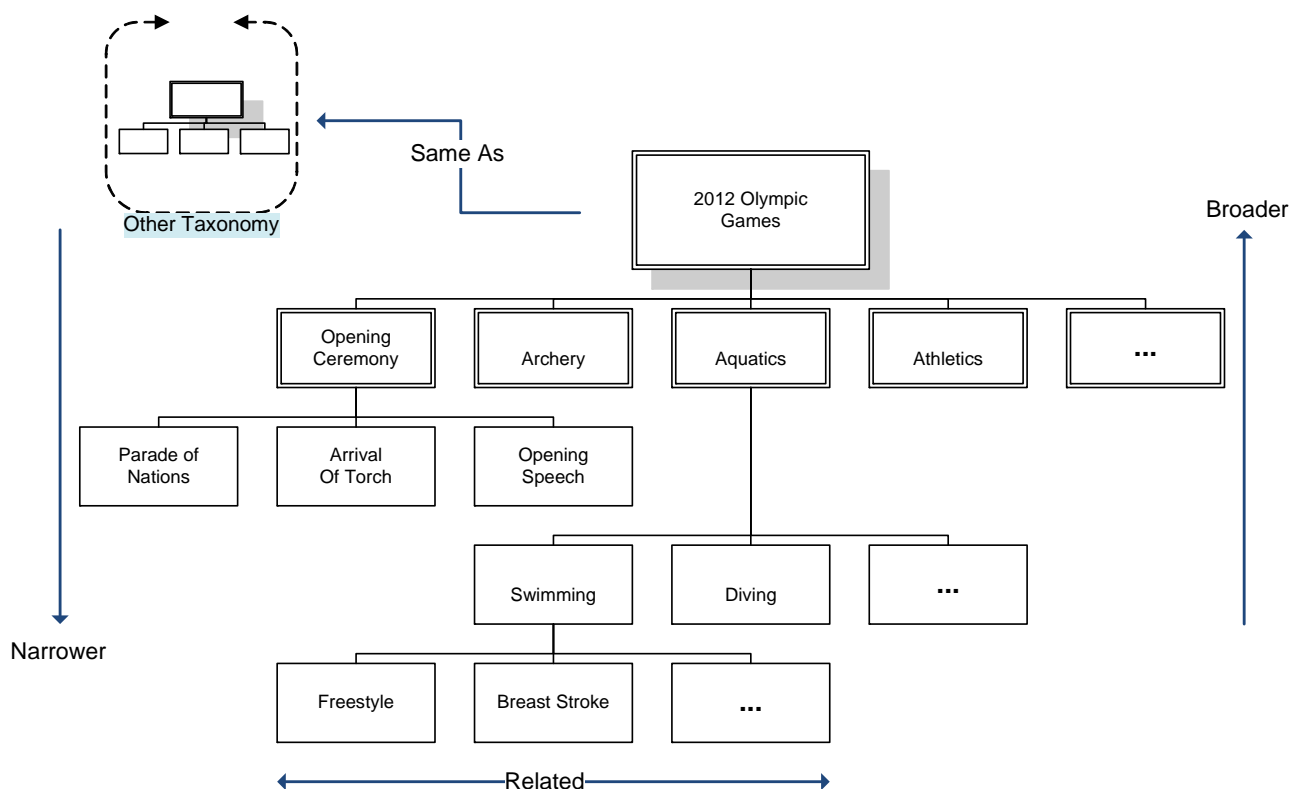
## 12.4.2 Event Relationships

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 may be to break down the 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 EventsML-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 22** below shows how a hierarchy of events may be created in this way and illustrates the meaning of <broader>, <narrower>, <related> and <sameAs> relationship properties.



**Figure 22: 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 (CCL)/Related Concept (PCL)	<p>Repeatable. The event may be part of another event, in which case this can be denoted by &lt;broader&gt;.</p> <p>We may use a literal value to identify the broader event that encompasses the current event:</p> <pre>&lt;broader   literal="int001"&gt;   &lt;name&gt;Treasury forecast of economic indicators&lt;/name&gt; &lt;/broader&gt;</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>&lt;broader   type="cpnat: event" qcode="events: 23456789"&gt;   &lt;name&gt;Treasury forecast of economic indicators&lt;/name&gt; &lt;/broader&gt;</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>&lt;narrower qcode="events: TR2009-34593"&gt;   &lt;name&gt;Minimum Lending Rate announcement press   conference&lt;/name&gt; &lt;/narrower&gt;</pre>
<related>	Flexible Property/Related Concept (PCL)	<p>Repeatable. May be used to denote a relationship to another concept or event, which is NOT an inherent characteristic (see Facet&gt;). 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>&lt;related   rel="irel: seeAlso"   type="cpnat: org"   qcode="org: fsauk"&gt;   &lt;name&gt;UK Financial Services Authority&lt;/name&gt; &lt;/related&gt;</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>&lt;sameAs qcode="coi: B0E30987"&gt;   &lt;name&gt;Monetary Policy Committee of the Bank of   Engl and&lt;/name&gt; &lt;/sameAs&gt;</pre>

### 12.4.3 Event Details Group

The event details group contains properties that are specific to EventsML-G2 and are wrapped by the `<eventDetails>` element. The first set of properties of Event Details is date-time information described in [12.4.3.1](#) below. The further properties are described in [12.4.3.2](#) below

#### 12.4.3.1 Dates and times

The `<dates>` wrapper **MUST** have a Start Date Time value amongst other properties detailed below, and **MUST** have EITHER an End Date Time OR Duration given for the event.

The IPTC recommends that when expressing time, the Time Zone is indicated (Z=UTC, or 0 offset)

Property	Type	Notes/Example
<code>&lt;start&gt;</code>	Approximate Date Time	<p>Mandatory, non-repeatable property has two optional attributes, <code>@approxstart</code> and <code>@approxend</code>.</p> <p>The value may be truncated, starting on the right (seconds) according to the precision required, but <b>MUST</b>, at minimum, have a year, for example:</p> <pre>&lt;start&gt;2009-06-12T12: 30: 00Z&lt;/start&gt;</pre> <p>or</p> <pre>&lt;start&gt;2009-06&lt;/start&gt;</pre> <p>The value of <code>&lt;start&gt;</code> expresses the <b>precise</b> date-time of the start of the event. With the information available, this might be a "best guess".</p> <p>By using <code>@approxstart</code> and <code>@approxend</code> 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 event itself, only the range of start date-times)</p> <p><code>@approxstart</code> indicates the start of the range. If used on its own, the end of the range of dates is the date-time value of <code>&lt;start&gt;</code></p> <p>For example, a possible start of an event on June 12, 2009, not before June 11, 2009, and no later than June 14 2009, would be expressed as:</p> <pre>&lt;start   approxstart=" 2009-06-11"   approxend=" 2009-06-14"&gt;   2009-06-12 &lt;/start&gt;</pre>
<code>&lt;end&gt;</code>	Approximate Date Time	<p>Non-repeatable element to indicate the 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 <code>&lt;start&gt;</code></p> <p>The <code>&lt;dates&gt;</code> wrapper <b>MUST</b> contain either an <code>&lt;end&gt;</code> date, or a <code>&lt;duration&gt;</code></p>
<code>&lt;duration&gt;</code>	XML Schema Duration	<p>Non-repeatable. The time period during which the event takes place is expressed in the form:</p> <p>PnYnMnDTnHnMnS</p> <p>P indicates the Period (required)</p> <p>nY = number of Years</p>



Property	Type	Notes/Example
		<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>Example:</p> <pre>&lt;duration&gt;PT3H&lt;/duration&gt;</pre> <p>The event will last for three hours. Note use of the "T" time separator even though no Date part is present.</p> <p>If the &lt;dates&gt; wrapper does not contain an &lt;end&gt; date, it <b>MUST</b> contain a &lt;duration&gt;</p>
<confirmation>	QCode	<p>Optional, non-repeatable. A QCode indicating whether the date-times for the event are confirmed or subject to possible change. The recommended IPTC NewsCode scheme currently has three possible values:</p> <pre>&lt;confirmation   qcode="edconf: bothApprox" /&gt;</pre> <p>Indicates that both start and end dates are currently approximate or undefined</p> <pre>&lt;confirmation   qcode="edconf: bothOk" /&gt;</pre> <p>Both start and end dates are confirmed</p> <pre>&lt;confirmation   qcode="edconf: startApprox" /&gt;</pre> <p>The start date is approximate but the end date is confirmed</p>

#### 12.4.3.1.1 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 to 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 specific a date, with an optional time part, or on several specific dates and times.</p> <pre>&lt;rdate&gt;2009-03-27T14: 00: 00Z&lt;/rdate&gt;</pre>



Property	Type	Notes/Example
		<code>&lt;rdate&gt;2009-04-03T16: 00: 00Z&lt;/rdate&gt;</code>
<code>&lt;rRule&gt;</code>	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> <p><code>&lt;rRule freq="MONTHLY" /&gt;</code></p> <p>The enumerated values of @freq are:</p> <ul style="list-style-type: none"> <li>❖ YEARLY</li> <li>❖ MONTHLY</li> <li>❖ DAILY</li> <li>❖ HOURLY</li> <li>❖ PER MINUTE</li> <li>❖ PER SECOND</li> </ul> <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> <p><code>&lt;rRule   freq="YEARLY"   interval="4" &gt;</code></p> <p>@until sets a Date with optional Time <b>after</b> which the recurrence rule expires:</p> <p><code>&lt;rRule   freq="MONTHLY"   until="2009-12-31" &gt;</code></p> <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> <p><code>&lt;rRule   freq="DAILY"   count="7" &gt;</code></p> <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 <a href="http://www.ietf.org/rfc/rfc2445.txt">http://www.ietf.org/rfc/rfc2445.txt</a></p> <p><code>&lt;start&gt;2009-01-11T8: 30: 00Z&lt;/start&gt; &lt;rRule   freq="YEARLY"   interval="2"   bymonth="1"   byday="su"   byhour="8 9"</code></p>



Property	Type	Notes/Example
		<pre>byminute="30" /&gt;</pre> <p>First, the interval="2" would be applied to freq="YEARLY" to arrive at "every other year". Then, bymonth="1" would be applied to arrive at "every January, every other year". Then, byday="SU" would be applied to arrive at "every Sunday in January, every other year".</p> <p>Then, byhour="8 9" (note that all multiple values are space separated) would be applied to arrive at "every Sunday in January at 8am and 9am, every other year". Then, byminute="30" would be applied to arrive at "every Sunday in January at 8:30am and 9:30am, every other year". Then, lacking information from rRule, the second is derived from &lt;start&gt;, to end up in "every Sunday in January at 8:30:00am and 9:30:00am, every other year".</p> <p>Similarly, if any of the @byminute, @byhour, @byday, @bymonthday or @bymonth rule part were missing, the appropriate minute, hour, day or month would have been retrieved from the &lt;start&gt; 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>&lt;rRule freq="MONTHLY" byday="MO TU WE TH FR" /&gt;</pre> <p>The same rule to specify the last working day of the month would be</p> <pre>&lt;rRule freq="MONTHLY" byday="MO TU WE TH FR" bypos="-1" /&gt;</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>&lt;rRule freq="WEEKLY" wkst="MO" /&gt;</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 &lt;exDate&gt;</p> <pre>&lt;rRule freq="MONTHLY" until="2009-12-31" /&gt;</pre>





Property	Type	Notes/Example
		<pre>&lt;exDate&gt;   2009-04-06 &lt;/exDate&gt;</pre>
<exRule>	Recurrence Rule	<p>Excluded Recurrence Rule. The same attributes as &lt;rRule&gt; 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>&lt;rRule   freq="WEEKLY"   until="2009-07-23" /&gt; &lt;rRule   freq="WEEKLY"   until="2009-12-24" /&gt; &lt;exRule   freq="WEEKLY"   until="2009-09-03" /&gt;</pre> <p>Note the order of the above statement: the &lt;rRule&gt; elements must come before &lt;exRule&gt;</p> <p>The meaning being expressed is:</p> <p>"The event occurs weekly until Dec 24, 2009 with a break from <b>after</b> July 23, 2009 until September 3, 2009.</p>

#### 12.4.3.2 Further Properties of Event Details

The event details group are properties that are specific to EventsML-G2 and are wrapped by the <eventDetails> wrapper 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><a href="http://cv.iptc.org/newscodes/eventoccurstatus/">http://cv.iptc.org/newscodes/eventoccurstatus/</a></p> <p>has values from "eos0" = "unplanned" through "eos5" = "certain to occur" indicating the provider's confidence of the status of the event.</p> <pre>&lt;occurStatus qcode="eocstat: eos5" /&gt;</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) &lt;name&gt; child element:</p> <pre>&lt;newsCoverageStatus qcode="ncstat: int"&gt;   &lt;name&gt;     Coverage Intended   &lt;/name&gt; &lt;/newsCoverageStatus&gt;</pre>



Property	Type	Notes/Example
<registration>	Block	<p>Optional, repeatable indicator of any registration details required for the event:</p> <pre>&lt;registration&gt;   Register online at &lt;br /&gt;   <a href="http://www.example.com/registration.aspx/">http://www.example.com/registration.aspx/</a> &lt;br /&gt; &lt;/registration&gt;</pre> <p>The property optionally takes a @role attribute. IPTC Registration Role NewsCode</p> <p><a href="http://cv.iptc.org/newscodes/eventregrole/">http://cv.iptc.org/newscodes/eventregrole/</a></p> <p>may be used, which currently has four values:</p> <ul style="list-style-type: none"> <li>❖ Exhibitor</li> <li>❖ Media</li> <li>❖ Public</li> <li>❖ Student</li> </ul> <pre>&lt;registration role="eregrol:exhibReg"&gt;   Exhibitors must register online at &lt;br /&gt;   <a href="http://www.example.com/exhibitor/register.aspx/">http://www.example.com/exhibitor/register.aspx/</a>   before May1, 2009&lt;br /&gt; &lt;/registration&gt; &lt;registration role="eregrol:pubReg"&gt;   The public may pre-register online at &lt;br /&gt;   <a href="http://www.example.com/public/register.aspx/">http://www.example.com/public/register.aspx/</a>   to receive a special bonus pack.&lt;br /&gt; &lt;/registration&gt;</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>&lt;access qcode="access: easy" /&gt;</pre>
<subject>	Flexible Property	<p>Optional, repeatable. The subject classification(s) of the event, for example, using the IPTC Subject NewsCodes:</p> <pre>&lt;subject   type="cpnat: abstract"   qcode="subj : 04000000"&gt;   &lt;name xml:lang="en-GB"&gt;     Economy, Business and Finance   &lt;/name&gt; &lt;/subject&gt; &lt;subject   type="cpnat: abstract"   qcode="subj : 04006000"&gt;   &lt;name xml:lang="en-GB"&gt;     Financial and Business Service   &lt;/name&gt; &lt;/subject&gt; &lt;subject</pre>



Property	Type	Notes/Example
		<pre> type="cpnat: abstract" qcode="subj : 04006002"&gt;   &lt;name xml:lang="en-GB"&gt;     Banki ng   &lt;/name&gt; &lt;/subj ect&gt; </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 &lt;name&gt; may be used.</p> <pre> &lt;l ocati on   type="cpnat: geoArea"   l i t e r a l ="bankofengl and"&gt;     &lt;name&gt;       The Bank of Engl and, Threadneedl e Street,       London, EC2R 8AH, UK     &lt;/name&gt; &lt;/l ocati on&gt; </pre> <p>At PCL, a rich Concept-style structure may be used. (See the G2 Specification document for details)</p>
<participant>	Flexible Property/ Flexible Party Property (PCL)	<p>Optional, repeatable, The people and/or organisations taking part in the event. The type of participant is identified by @type and a QCode. The following example indicates a person, an organisation would be indicated (using the IPTC NewsCode) by type="cpnat:org"</p> <pre> &lt;parti ci pant   type="cpnat: person"   qcode="pers: 32965"&gt;     &lt;name xml:lang="en-GB"&gt;       Mervyn Ki ng     &lt;/name&gt; &lt;/parti ci pant&gt; </pre> <p>An IPTC Event Participant Role NewsCode is available:  <a href="http://cv.iptc.org/newscodes/eventparticipantrole/">http://cv.iptc.org/newscodes/eventparticipantrole/</a>  that holds roles such as "moderator", "director", "presenter"</p>
<participationRequirement>	Flexible Property	<p>Optional, repeatable element for expressing any required conditions for participation in, or attendance at, the event. Either a literal or QCode value may be used.</p> <pre> &lt;parti ci pati onRequi rement   l i t e r a l ="accredi ted"&gt;     &lt;name&gt;       Accredi tati on requi red     &lt;/name&gt; &lt;/parti ci pant&gt; </pre>
<organiser>	Flexible Property/ Flexible Party	<p>Optional, repeatable. Describes the organiser of the event.</p> <pre> &lt;organi ser   type="cpnat: org"   l i t e r a l ="i p t c"&gt;     &lt;name xml:lang="en-GB"&gt; </pre>



Property	Type	Notes/Example
	Property (PCL)	<pre> International Press Telecommunications Council &lt;/name&gt; &lt;name xml-lang="fr"&gt; Comité International de Télécommunications de Presse &lt;/name&gt; &lt;/organiser&gt; </pre> <p>The IPTC Event Organiser Role NewsCode, viewable at <a href="http://cv.iptc.org/newscodes/eventorganiserrole/">http://cv.iptc.org/newscodes/eventorganiserrole/</a> 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 <b>12.4.3.2.1</b> below
<language>	-	<p>Optional, repeatable element describes the language(s) associated with the event using @tag with values that must conform to the IETF's BCP 47. An optional child element &lt;name&gt; may be added.</p> <pre> &lt;language tag="en"&gt;   &lt;name&gt;English&lt;/name&gt; &lt;/language&gt; </pre>
<newsCoverage>	Complex (CCL) XML Mixed (PCL)	<p>Optional, repeatable property indicates the journalistic endeavour and content that is planned in response to the event. At CCL, the element holds a natural language description of planned coverage in a repeatable &lt;edNote&gt; child element:</p> <pre> &lt;newsCoverage&gt;   &lt;edNote&gt;     Our staff reporter will be at the event from 9.30am and is expected to file a 250 Word bulletin by 11.30am and a 500 word lead by 12.30pm. All times GMT.   &lt;/edNote&gt;   &lt;edNote&gt;     There is a photocall at 11.30am. Our Photographer Will file a selection of landscape and portrait pictures by 12noon. All times GMT.   &lt;/edNote&gt; &lt;/newsCoverage&gt; </pre> <p>At PCL a larger set of properties may be used to create more detailed and machine-readable coverage information. See <b>12.4.3.2.2</b> below for more details of available properties.</p>

#### 12.4.3.2.1 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>&lt;email role="addressrole: press"&gt;   office@iptc.org &lt;/email&gt;</pre>
<im>	Electronic Address	<pre>&lt;im role="imsrvc: reuters"&gt;   j.doe@iptc.org@reuters.net &lt;/im&gt;</pre>
<phone>	Electronic Address	Repeatable. Phone numbers should have a @role if necessary distinguish between different numbers: <pre>&lt;phone role="phrole: public"&gt;   1-123-456-7899 &lt;/phone&gt; &lt;phone role="phrole: press"&gt;   1-123-456-7898 &lt;/phone&gt;</pre>
<fax>	Electronic Address	Fax number(s), as described for <phone> above.
<web>	IRI	<pre>&lt;web role="webrole: event"&gt;   www.iptc.org/springmeeting.html &lt;/web&gt;</pre>
<address>	Address	See 9.6.1.4 below for details of Address properties. The Address may have a @role to denote the type of address it contains (e.g. office, personal) and may be repeated as required to express each address @role.
<note>	Block	Any other contact-related information, such as “Office closed for lunch from 12.30pm for three hours”

#### 12.4.3.2.1.1 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="addressrole: registered">
```

```
<line>20 Garri ck Street</line>
<locality literal="London" />
<country qcode="ISOCountryCode: UK">
  <name xml:lang="en">United Kingdom</name>
</country>
<postalCode>WC2E 9BT</postalCode>
</address>
```

#### 12.4.3.2.2 News Coverage at Power Conformance Level

At CCL News Coverage information is expressed as a Block type element, with some mark-up permitted. At PCL, more extensive, detailed information may be given as discussed below.

The advantages of using PCL capabilities for News Coverage are that more machine-readable information can be used to populate customer's resource management applications, and that detailed descriptive metadata may be associated with each announced <newsCoverage> structure, ready to be inherited by the arriving content, thus speeding up news handling and potentially increasing consistency and quality.

Property	Type	Notes/Example
<newsCoverage>	Mixed	<p>Available (optional) attributes are:</p> <p>@role to indicate what part this plays in the announced coverage, using a QCode</p> <p>@id is a local identifier for this News Coverage element. The scope of the id is LOCAL (only within the document)</p> <p>@modified is the Date (and optionally the time) when the property was modified.</p> <pre>&lt;newsCoverage modified="2008-01-26T13:19:11Z"&gt;   &lt;itemClass qcode="ni nat: text"&gt;     &lt;name&gt;text&lt;/name&gt;   &lt;/itemClass&gt;   &lt;scheduled&gt;     2009-10-16T17:00:00+02:00   &lt;/scheduled&gt;   &lt;edNote&gt;250 words&lt;/edNote&gt;   &lt;genre qcode="mygenre: leadtext"&gt;     &lt;name&gt;Main text story&lt;/name&gt;   &lt;/genre&gt; &lt;/newsCoverage&gt;</pre>
<g2ContentType>	String	<p>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.</p> <pre>&lt;g2ContentType&gt;   Appl ication/vnd. iptc. g2. newsi tem+xml &lt;/g2ContentType&gt;</pre>
<itemClass>	QCode	<p>Optional, non-repeatable element indicates the type of content to be delivered, using a Controlled Vocabulary, such as the IPTC News Item Nature NewsCode. Since the example will show a text article, the Item Class is "text"</p> <pre>&lt;itemClass qcode="ni nat: text" /&gt;</pre>
<assignedTo>	Flex1 Party Property	<p>The Flex1 Party Property Type extends the Flex Party Property Type by allowing a @role attribute to be a space separated list of QCodes.</p>



Property	Type	Notes/Example
		<p>&lt;assignedTo&gt; 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 &lt;personDetails&gt; or &lt;organisationDetails&gt;, and properties from the Concept Definitions Group, and Concepts Relationships Group (see <a href="#">Concepts and Concept Items</a> for details).</p> <p>The example shows the details of a person assigned to create the content:</p> <pre>&lt;assignedTo   role="editor"   type="cpnat: person"   qcode="pers: 54321" &gt;   &lt;name&gt;Sue Fine&lt;/name&gt;   &lt;personDetails&gt;     &lt;contactInfo&gt;       &lt;phone&gt;1-418-4567&lt;/phone&gt;       &lt;email&gt;editor@iptc.org&lt;/email&gt;     &lt;/contactInfo&gt;   &lt;/personDetails&gt; &lt;/assignedTo&gt;</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 &lt;by&gt; property (see below)</p>
<scheduled>	Approx Date Time	<p>Optional, non-repeatable. Indicates the scheduled time of delivery, and may be truncated if the precise date and time is not known. For example, if the content is scheduled to arrive at some unspecified time on a day, the value would be, for example:</p> <pre>&lt;scheduled&gt;2009-10-16&lt;/scheduled&gt;</pre>
<service>	Qualified Property	<p>Optional, repeatable. The editorial service to which the content has been assigned by the provider and on which the receiver should expect to receive the planned content.</p> <pre>&lt;service qcode="srv: intwire"&gt;   &lt;name&gt;International Wire Service&lt;/name&gt; &lt;/service&gt;</pre>
<edNote>	Block	<p>Optional, repeatable. Editorial note giving natural language information on the planned coverage</p> <pre>&lt;edNote&gt;Additional media release expected&lt;/edNote&gt;</pre>
<by>	Label	<p>Optional, repeatable. Natural language author/creator information</p> <pre>&lt;by&gt;By Sue Fine&lt;/by&gt;</pre>
<dateline>	Label	<p>Optional, repeatable. Natural language information traditionally placed at the start of a text by some news agencies, indicating</p>





Property	Type	Notes/Example
		<p>the place and time that the content was created</p> <pre>&lt;date line&gt;   Tel Aviv, January 27, 2009 (Reuters) &lt;/date line&gt;</pre>
<description>	Block	<p>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"</p> <pre>&lt;description role="drole: summary"&gt;   Sue Fine will report on the proceedings from the   NewsCodes Working Party &lt;/description&gt;</pre>
<genre>	Flex1 Concept Property	<p>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.</p> <p>Child elements may be any from the Concept Definition Group and the Concept Relationships Group (see <a href="#">Concepts and Concept Items</a>)</p> <pre>&lt;genre qcode="mygenre: main"&gt;   &lt;name&gt;Main article &lt;/name&gt; &lt;/genre&gt;</pre>
<headline>	Label	<p>Optional, repeatable. Headline that will apply to the content.</p> <pre>&lt;headline&gt;NewsCodes Working Party&lt;/headline&gt;</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 &lt;name&gt;</p> <pre>&lt;language tag="en-GB"&gt;   &lt;name&gt;UK English&lt;/name&gt; &lt;/language&gt;</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>&lt;slugline separator=" - "&gt;US-AUTO-BAILOUT&lt;/slugline&gt;</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 <a href="#">Concepts and Concept Items</a>)</p>



Property	Type	Notes/Example
		<pre> &lt;subject qcode="subj : 04010000"&gt;   &lt;name&gt;medi a&lt;/name&gt; &lt;/subject&gt; &lt;subject qcode="subj : 04010004"&gt;   &lt;name&gt;news agency&lt;/name&gt; &lt;/subject&gt; &lt;subject qcode="subj : 13022000"&gt;   &lt;name&gt;I T/computer sci ences&lt;/name&gt; &lt;/subject&gt;  Or  &lt;subject qcode="subj : 04010004"&gt;   &lt;name&gt;news agency&lt;/name&gt;   &lt;broadener qcode="subj : 04010000"&gt;     &lt;name&gt;Medi a&lt;/name&gt;   &lt;/broadener&gt; &lt;/subject&gt; &lt;subject qcode="subj : 13022000"&gt;   &lt;name&gt;I T/computer sci ences&lt;/name&gt; &lt;/subject&gt; </pre>

## 12.5 Event Use Cases

The following use cases illustrate the three methods of conveying event information: as the content of a G2 News Item, as a G2 Concept, and as part of a set of Concepts in a Knowledge Item.

### 12.5.1 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 at the start of day listing the events of the day. In many cases, these were provided as a text story, with minimal mark-up.

Using EventsML-G2 properties, these can be conveyed as a list of events capable of being machine-processed, enabling receivers to, for example, populate an in-house calendar, display on a company Intranet, or format as a user-friendly document.

One of the available child elements of a News Item <contentSet> is the <inlineXML> wrapper element, which can convey any valid XML content – in this case an <events> wrapper containing one or more instances of <event>, each of which is a separate piece of self-contained event information.

Using this method, we can convey the event listing inside NewsML-G2. The document is an instance of the NAR “anyItem” class, with the <newsItem> root element 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/
  NewsML-G2_2.2-spec-NewsItem-Power.xsd
  guid="urn:newsml:iptc.org:20090122:qqwpi ruew4711"
  version="1" standard="NewsML-G2" standardversion="2.2" conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_6.xml" />

```

The <itemMeta> block holds management information about the NewsML-G2 document (NOT the events conveyed by the document). The Item Class has a value from the IPTC News Item Nature NewsCode of “composite”, together with Provider, a Time Stamp and Publishing Status

```

<itemMeta>
  <itemClass qcode="ni nat: composite"/>
  <provider literal="IPTC"/>
  <versionCreated>2009-01-22T12:00:00Z</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 a child elements chosen from either <inlineXML>, <inlineData> or <remoteContent>. In this case we will be using <inlineXML> to wrap an <events> structure. (Event information abridged for ease of reading)

#### LISTING 15 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/
    NewsML-G2_2.2-spec-NewsItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090122:qqwpi ruuew4711"
  version="1" standard="NewsML-G2" standardversion="2.2" conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_6.xml" />
  <itemMeta>
    <itemClass qcode="ni nat: composite" />
    <provider literal="IPTC" />
    <versionCreated>2009-01-22T12:00:00Z
    </versionCreated>
    <pubStatus qcode="stat: usable" />
  </itemMeta>
  <contentSet>
    <inlineXML>
      <events>
        <!-- X -->
        <!-- FIRST EVENT! -->
        <!-- X -->
        <event>
          <name>IPTC Autumn Meeting</name>
          <eventDetails>
            <dates>
              <start>2009-10-15T09:00:00+02:00</start>
              <duration>P3D</duration>
            </dates>
            <location literal="SASRadissonAlcronHotelPrague">
              <name>SAS Radisson Alcron Hotel, Prague, Czech Republic</name>
              <facet rel="frel: venue type" qcode="ventyp: hotel" />
              <POIDetails>
                <position latitude="50.0809" longitude="14.428096" />
                <contactInfo>
                  <web>http://www.sasradisson.com</web>
                </contactInfo>
              </POIDetails>
            </location>
          </eventDetails>
        </event>
        <!-- X -->
        <!-- SECOND EVENT! -->
        <!-- X -->
        <event>
          <name>IPTC News Content Working Party session at the IPTC
            Autumn Meeting</name>
          <eventDetails>
            <dates>
              <start>2009-10-15T09:30:00+02:00</start>
              <end>2009-10-15T18:00:00+02:00</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.
          </definition>
        </event>
      </events>
    </inlineXML>
  </contentSet>
</newsItem>
```

```

        <br />
      </definition>
      <facet rel="frel:facility" qcode="facIncd:Food" />
      <facet rel="frel:facility" qcode="facIncd:AirConditioning" />
      <eventDetails>
        <dates>
          <start>2007-08-02T19:30:00+01:00</start>
          <end>2007-08-31</end>
          <rRule freq="DAILY" byday="TH FR SA" />
        </dates>
      </eventDetails>
    </event>
  </events>
</inlineXML>
</contentSet>
</newsItem>

```

## 12.5.2 Event as a Concept

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. But many events need to be persisted in a knowledge base so that they can be updated when necessary, and referenced many times over a period of time.

Events conveyed in News Items cannot easily be managed in this way. If an event needs to be managed over time, it must be conveyed as a Concept, either in a Concept Item or Knowledge Item, with a Concept ID that enables it to be stored with an unambiguous identifier.

Some providers are developing and trialling an instantaneous “push” events service, where customers receive a stream of event information as G2 Concept Items which directly populates the customer’s own content management system.

In this case, EventsML-G2 is being used as a bridge connecting 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.

To show how this is done, we will re-cast one of the example events in [12.5.1](#) above as a Concept carried in a Concept Item. [Figure 23](#) below shows how the essential event information sits unchanged within the different wrapper elements of News Item and Concept Item

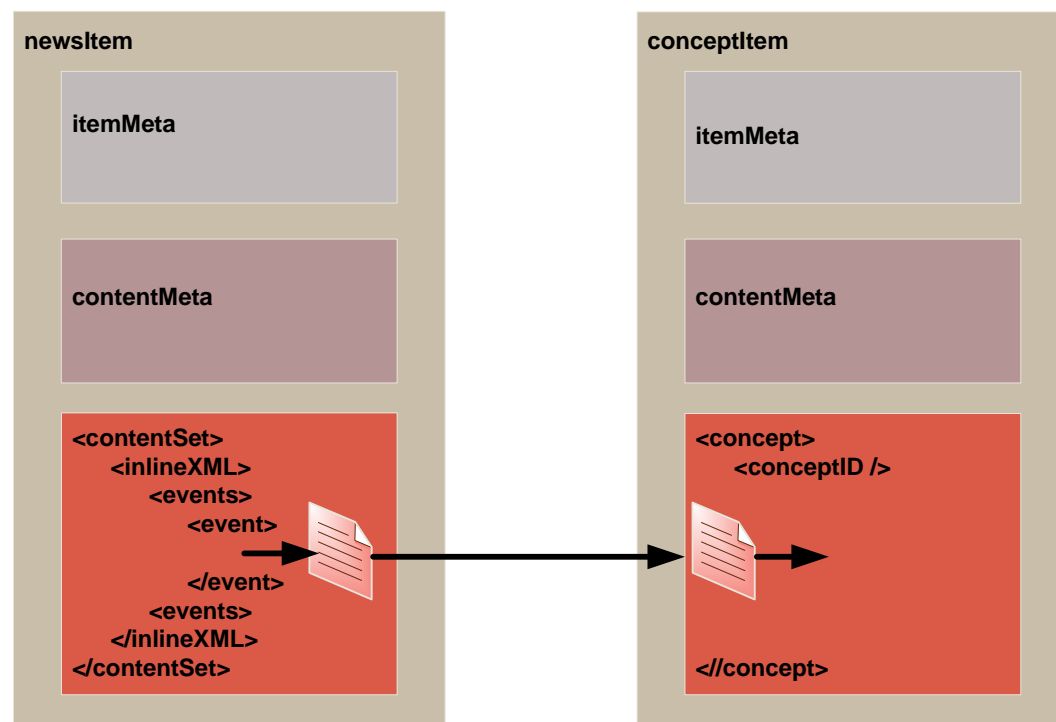


Figure 23: An event represented in a News Item or Concept Item uses a common structure



The top level element of the document is changed to <conceptItem> and some of its properties are also changed to reflect that the XML Schema is different (G2 Standards use a modular set of XML schemas).

All G2 documents must be uniquely identified by a GUID. By re-sending event information with the same GUID and an updated version number, the receiver can be advised to add or replace previously-sent information.

Since all concepts must be identified by a ConceptId QCode, a reference to the provider's catalog **MUST** be included (or a catalog statement giving the scheme URI)

```
<?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/
    EventsML-G2_1.1-spec-ConceptItem-Power.xsd"
  guid="urn:newsmil:iptc.org:20090126:qqwpi ruuew4711"
  version="1"
  standard="EventsML-G2"
  standardversion="1.1"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <catalogRef
    href="http://www.example.com/events/event-catalog.xml" />
/>
```

Top level element

XML Schema change

GUID for the document and version

Standard: EventsML-G2

Provider's catalog

The <itemMeta> and <contentMeta> (if applicable) blocks 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 to tell the receiver that the concepts make up a controlled vocabulary or taxonomy.)

```
<itemMeta>
  <itemClass qcode="cinat:concept" />
  <provider literal="IPTC" />
  <versionCreated>2009-01-26T14:00:00Z
</versionCreated>
  <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
  <urgency>5</urgency>
  <contentCreated>2009-01-26T12:15:00Z</contentCreated>
  <contentModified>2009-01-26T13:35:00Z</contentModified>
</contentMeta>
```

Concept Item nature

Instead of <contentSet>, the child element of a Concept Item is <concept>. Within each Concept, we need a <conceptId>, a QCode value that will provide a way for other events or concepts to reference the Concept being conveyed. The type of concept being conveyed is also expressed using the <type> property with a QCode to tell the receiver that it contains Event information, using the IPTC "Nature of Concept" NewsCode. (Values include "event", "organisation", "person") The CV may be viewed at [www.cv.iptc.org/newcodes/cpnature/](http://www.cv.iptc.org/newcodes/cpnature/).

```
<concept>
  <conceptId created="2009-01-26T12:00:00Z" qcode="event:1234567" />
  <type qcode="cpnat:event"></type>
```

This is an Event

The other change from News Item to Concept Item is to remove the <events> and <event> wrapper elements, to place the event information, such as <name> and <eventDetails> directly within the Concept.

```
<concept>
  <conceptId created="2009-01-26T12:00:00Z" qcode="event:1234567" />
  <type qcode="cpnat:event" />
  <name>IPTC Autumn Meeting</name>
  <eventDetails>
    <dates>
      <start>2009-10-15T09:00:00+02:00</start>
      <duration>P3D</duration>
    </dates>
    . . . . .
```



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/
    EventsML-G2_1.1-spec-ConceptItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090122:qqwpi ruuew4711"
  version="1"
  standard="EventsML-G2"
  standardversion="1.1"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_6.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="ci nat: concept" />
    <provider literal="IPTC" />
    <versionCreated>2009-01-22T13:45:00Z
    </versionCreated>
    <pubStatus qcode="stat: usable" />
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2009-01-26T12:15:00Z</contentCreated>
    <contentModified>2009-01-26T13:35:00Z</contentModified>
  </contentMeta>
  <concept>
    <conceptId created="2009-01-26T12:00:00Z" qcode="event: 1234567" />
    <type qcode="cpnat: event" />
    <name>IPTC Autumn Meeting</name>
    <eventDetails>
      <dates>
        <start>2009-10-15T09:00:00+02:00</start>
        <duration>P3D</duration>
      </dates>
      <location literal="SASRadissonAlcronHotel Prague">
        <name>SAS Radisson Alcron Hotel, Prague, Czech Republic</name>
        <facet rel="frel: venue type" qcode="ventyp: hotel" />
        <POIDetails>
          <position latitude="50.0809" longitude="14.428096" />
          <contactInfo>
            <web>http://www.sasradisson.com</web>
          </contactInfo>
        </POIDetails>
      </location>
    </eventDetails>
  </concept>
</conceptItem>
```

After some passage of time, the provider may have organised news coverage of this event, and will send an update to the customer. 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/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    EventsML-G2_1.1-spec-ConceptItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090126:qqwpi ruuew4711" ← same GUID
  version="2" ← new version number
  standard="EventsML-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="ci nat: concept" />
  <provider literal="IPTC" />
  <versionCreated>2009-01-26T14:45:00Z ← Item timestamp
```





```

</versionCreated>
<pubStatus qcode="stat:usable" />
<edNote>Updated to indicate our coverage plans</edNote>
<signal qcode="eventproc:replace">
  <name>Replace existing</name>
</signal>
</itemMeta>
<contentMeta>
<urgency>5</urgency>
<contentCreated>2009-01-26T12: 15: 00Z</contentCreated>
<contentModified>2009-01-26T14: 35: 00Z</contentModified>

```

Editorial note  
Signal to processor  
Natural language name of instruction  
Concept timestamp

In the following listing, the complete event information is re-sent, with a processing instruction that the previous version should be replaced. Two <newsCoverage> blocks tell the receiver of the content that has been planned in response to the event, the type of content that is planned, and the scheduled time of arrival.

As this document is declared to be at the Power Conformance Level, there is more scope for giving the News Coverage in detail. At CCL, news coverage information is expressed in natural-language block elements.

#### LISTING 17 Update to the Previously-Sent Event

```

<?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/
    EventsML-G2_1.1-spec-ConceptItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090126:qqwpi ruuew4711"
  version="2"
  standard="EventsML-G2"
  standardversion="1.1"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_6.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
    <versionCreated>2009-01-26T14: 45: 00Z</versionCreated>
    <pubStatus qcode="stat:usable" />
    <edNote>Updated to indicate our coverage plans</edNote>
    <signal qcode="eventproc:replace">
      <name>Replace existing</name>
    </signal>
  </itemMeta>
  <contentMeta>
    <urgency>5</urgency>
    <contentCreated>2009-01-26T12: 15: 00Z</contentCreated>
    <contentModified>2009-01-26T14: 35: 00Z</contentModified>
  </contentMeta>
  <concept>
    <conceptId created="2009-01-26T12: 00: 00Z" qcode="event: 1234567" />
    <type qcode="cpnat:event"></type>
    <name>IPTC Autumn Meeting</name>
    <eventDetails>
      <dates>
        <start>2009-10-15T09: 00: 00+02: 00</start>
        <duration>P3D</duration>
      </dates>
      <location literal="SASRadissonAlcronHotel Prague">
        <name>SAS Radisson Alcron Hotel, Prague, Czech Republic</name>
        <facet rel="frel:venue type" qcode="ventyp:hotel" />
        <POIDetails>
          <position latitude="50.0809" longitude="14.428096" />
          <contactInfo>
            <web>http://www.sasradisson.com</web>
          </contactInfo>
        </POIDetails>
      </location>
      <newsCoverage id="ID_1234568" modified="2008-01-26T13: 19: 11Z">
        <itemClass qcode="ninat:text" />
        <scheduled>2009-10-16T17: 00: 00+02: 00</scheduled>
        <edNote>250 words</edNote>
        <genre qcode="mygenre:leadtext">
          <name>Main text story</name>
        </genre>
      </newsCoverage>
    </concept>
  </eventDetails>
</concept>

```





```

        </genre>
    </newsCoverage>
    <newsCoverage id="ID_1234569" modified="2009-01-26T15:19:11Z">
        <itemClass qcode="inat:picture" />
        <scheduled>2009-10-16T13:00:00+02:00</scheduled>
    </newsCoverage>
</eventDetails>
</concept>
</conceptItem>

```

### 12.5.3 Multiple Event Concepts in a Knowledge Item

A third method of conveying event information is as a set of concepts in a Knowledge Item, which shares a common structure with other G2 Items, but conveys a Concept Set, containing 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, as do other G2 Items.

```

<?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/
    EventsML-G2_1.1-spec-KnowledgeItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090126:qqwpi ruuew4712"
  version="1"
  standard="EventsML-G2"
  standardversion="1.1"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_6.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />

```

Top Level element

XML Schema change 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 Descriptive Metadata elements that may be used with a Knowledge item are <subject> and <description>.

```

<itemMeta>
  <itemClass qcode="inat:concept" />
  <provider literal="IPTC" />
  <versionCreated>2009-01-26T12:00:00Z</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>

```

item conveys concepts

doc version timestamp

Item is useable

Content timestamps

Subject codes of the events

The two events in the listing are related, with the relationship indicated by the second event using the <broaden> 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 -->
  </concept>
  <conceptId
    created="2000-10-30T12:00:00+00:00"
    qcode="event:1234567"

```

Concept ID



```

/>
<name>IPTC Autumn Meeting</name>
<eventDetails>
.....

```

Second event:

```

<concept>
  <!-- SECOND EVENT! -->
  <!-- x -->
  <conceptId
    created="2000-10-30T12:00:00+00:00"
    qcode="event:91011123"
  />
  <name>IPTC News Content Working Party session at the IPTC Autumn
    Meeting</name>
  <broadener
    type="cinat:event"
    qcode="event:1234567">
    <name>IPTC Autumn Meeting</name>
  </broadener>
</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"?>
<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/
    EventsML-G2_1.1-spec-KnowledgeItem-Power.xsd"
  guid="urn:newsml:iptc.org:20090126:qqwpi ruuew4712"
  version="1"
  standard="EventsML-G2"
  standardversion="1.1"
  conformance="power"
  xml:lang="en">
  <catalogRef href="http://www.iptc.org/std/catalog/IPTC-G2-Standards_6.xml" />
  <catalogRef href="http://www.example.com/events/event-catalog.xml" />
  <itemMeta>
    <itemClass qcode="cinat:concept" />
    <provider literal="IPTC" />
    <versionCreated>2009-01-26T12:00:00Z</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="2000-10-30T12:00:00+00:00" qcode="event:1234567" />
      <name>IPTC Autumn Meeting</name>
      <eventDetails>
        <dates>
          <start>2009-10-15T09:00:00+02:00</start>
          <duration>P3D</duration>
        </dates>
        <registration>Registration with the IPTC office is required.
          A <a href="http://www.iptc.org/ThisIsNotSerious.php">
            web form</a> may be used until 11 September 2009
        </registration>
      </eventDetails>
    </concept>
  </conceptSet>

```



```

    </regi strati on>
    <parti ci pati onRequi rement>
      <name>Membershi p</name>
      <defi ni ti on>Only members of the IPTC and their invited
        guests may attend
      </defi ni ti on>
    </parti ci pati onRequi rement>
    <accessStatus qcode="accst: easy" />
    <language tag="en" />
    <organi ser Li teral ="IPTC" role="orgrole: mai nOrgani ser">
      <name>International Press Tel ecommuni cati ons Council</name>
      <organi sati onDetail s>
        <founded>1965</founded>
      </organi sati onDetail s>
    </organi ser>
    <contactI nfo>
      <email >mdi rector@i pct. org</email >
      <note>Mi chael Stei dl , Managi ng Di rector</note>
      <web>http: //www. i ptc. org</web>
    </contactI nfo>
    <locati on Li teral ="SASRadi ssonAl cronHotel Prague">
      <name>SAS Radi sson Al cron Hotel , Prague, Czech Republ ic</name>
      <facet rel ="frel: venuetype" qcode="ventyp: hotel " />
      <POI Detail s>
        <posi ti on La ti tude="50. 0809" l ongi tude="14. 428096" />
        <contactI nfo>
          <web>http: //www. sasradi sson. com</web>
        </contactI nfo>
      </POI Detail s>
    </locati on>
    <parti ci pant Li teral ="StephaneGueri l l ot">
      <name>Stéphane Guéri l l ot</name>
      <defi ni ti on role="drol: jobti tle">IPTC Chai rman</defi ni ti on>
    </parti ci pant>
    <parti ci pant Li teral ="Mi chael Stei dl ">
      <name>Mi chael Stei dl </name>
      <defi ni ti on role="drol: jobti tle">Managi ng Di rector</defi ni ti on>
    </parti ci pant>
    <newsCoverage>
      <edNote>Expect an IPTC medi a rel ease on Tuesday, 16 October 2009</edNote>
    </newsCoverage>
  </eventDetail s>
</concept>
<concept>
  <!-- SECOND EVENT! -->
  <!-- X -->
  <conceptId created="2000-10-30T12:00:00+00:00" qcode="event: 91011123" />
  <name>IPTC News Content Working Party session at the IPTC Autumn
    Meeting</name>
  <broaden type="cpnat: event" qcode="event: 1234567">
    <name>IPTC Autumn Meeting</name>
  </broaden>
  <eventDetail s>
    <dates>
      <start>2009-10-15T09:30:00+02:00</start>
      <end>2009-10-15T18:00:00+02:00</end>
    </dates>
    <parti ci pati onRequi rement>
      <name>Membershi p</name>
      <defi ni ti on>Only members of the IPTC and their invited
        guests may attend
      </defi ni ti on>
    </parti ci pati onRequi rement>
    <accessStatus qcode="accst: easy" />
    <language tag="en" />
    <parti ci pant Li teral ="DarkoGul i ja">
      <name>Darko Gul i ja</name>
      <defi ni ti on role="drol: jobti tle">NCT WP Chai rman</defi ni ti on>
    </parti ci pant>
    <parti ci pant Li teral ="Mi chael Stei dl ">
      <name>Mi chael Stei dl </name>
      <defi ni ti on role="drol: jobti tle">Managi ng Di rector</defi ni ti on>
    </parti ci pant>
  </eventDetail s>
</concept>
</conceptSet>
</knowl edgel tem>

```



## 13 SportsML-G2

### 13.1 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.

### 13.2 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 using G2.

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

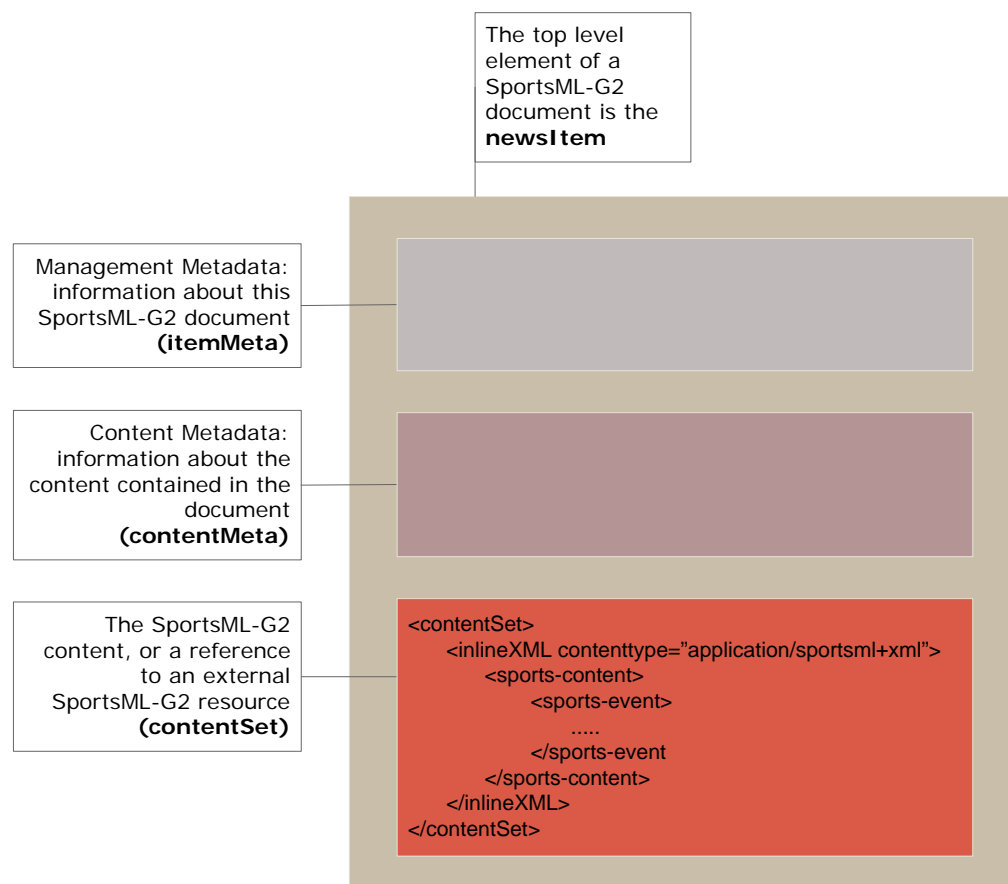
Finally, by building expertise around a single, extensible standard, information owners, providers, and consumers can benefit from reduced costs, greater reliability and faster time-to-market for compelling applications based on the rich variety of sports statistics and content that is increasingly available.

### 13.3 Structure

SportsML-G2 content is conveyed within existing 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 24](#).

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 24: 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.

### 13.3.1 XML Schema and Catalogs

The schema used by the G2 <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 "2.2". 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](http://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_6.xml" />
<catalog>
  <scheme alias="fixture" uri="http://xmlteam.org/newscodes/fixture#" />
</catalog>
```

### 13.3.2 Item Metadata

A typical <itemMeta> block<sup>12</sup> may include the following:

```
<itemMeta>
  <itemClass qcode="ni nat: text" />
  <provider qcode="web: www.xml team. com" />
  <versionCreated>2009-02-02T11: 17: 00Z</versionCreated>
  <pubStatus qcode="stat: usable" />
  <fileName>xt. 5932656-previous.xml</fileName> ← recommended filename
</itemMeta>                                     for the item
```

### 13.3.3 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 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="ci ty: Phi I adel phi a"> ← where created
    <broader qcode="reg: Pennsyl vani a" />
    <broader qcode="cntry: USA" />
  </located>
  <creator qcode="web: sportsnetwork. com"> ← created by
    <name>The Sports Network</name>
  </creator>
  <altId type="i dtype: tsn-i d" ← alternative ids
    id="sportsnetwork. com-5932656" />      with @type
  <altId type="i dtype: revisi on-i d"
    id="l. ml b. com-2009-e. 19358-pre-event-coverage-sportsnetwork. com" />
  <genre ← content genre
    type="spcpnat: spfi xt"
    qcode="spfi xt: pi tcher-previous">
      <name xml: lang="en-US">Game Pi tcher Previ ew</name>
      <broader qcode="dcl ass: event-summary" />
    </genre>
    <genre type="spcpnat: dcl ass"
      qcode="dcl ass: event-summary" />
    <genre type="xts-genre: tsn-fi xt看ure"
      qcode="tsn-fi xt看ure: ml bprevi ewxml " />
  </genre>
  <language tag="en-US" />
  <subject qcode="subj: 15000000"> ← about the content
    <name xml: lang="en-US">sport</name>      (subject codes)
  </subject>
  <subject qcode="subj: 15007000">
    <name xml: lang="en-US">Basebal l</name>
    <broader qcode="subj: 15000000" />
  </subject>
```

<sup>12</sup> 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](http://www.iptc.org). This shows a comparison between pre-G2 structures and elements, and their G2 equivalents.





```
</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
<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="fssld: 45679"/> ← "same as" player id
</subject>
<subject type="cpnat: person"
  qcode="person: l . ml b. com-p. 123">
  <name role="nrol: full">Freddie Garcia</name>
  <sameAs qcode="fssld: 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>
```

### 13.3.4 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 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

#### 13.3.4.1 Schedules or Fixtures <schedule>

A series of games, usually grouped by date. Also known as a fixture list.





#### 13.3.4.2 Sports Event <sports-event>

Contains metadata and data about a sporting competition – a sporting content that generally ends with a winner

#### 13.3.4.3 Standing <standing>

A series of team or individual records; facts about a team or player's performances during a match, competition, season or career,

#### 13.3.4.4 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.

#### 13.3.4.5 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

### 13.3.5 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.

#### 13.3.5.1 Event Metadata <event-metadata>

The contents of <event-metadata> are sports-specific: for example, for rugby, the time added to the match by the officials may be indicated.

Additionally, there are properties for recording the Event Sponsor and details of the Event Venue, The following information might apply to the Australian Open men's tennis final:

```
<event-metadata date-coverage-type="event"
  event-key="event: l.atp.com-2009-e.19358" event-status="mid-event"
  start-date-time="2009-01-28T19:05:00+11:00">
  <event-sponsor
    type="main"
    name="Kia Motors">
  </event-sponsor>
  <site>
    <site-metadata capacity="15,000" surface="Plexicushion"
      home-page-url="www.australianopen.com" site-key="site: aus209">
      <name>Rod Laver Arena, Melbourne Park</name>
    </site-metadata>
    <site-stats attendance="13,879" temperature="38"
      temperature-units="Centi grade" />
  </site>
</event-metadata>
```

There is a choice of more than 40 attributes of <event-metadata> to record detailed information about an event, its timing, the venue, predicted weather – even its suitability for record-breaking on the day.

#### 13.3.5.2 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.

### 13.3.5.3 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.mlb.com-t.19" alignment="home">
    <name full="Philadelphia Phillies" role="nrol: full">
      Philadelphia Phillies
    </name>
  </team-metadata>
  <player>
    <player-metadata position-event="1" player-key="player: l.mlb.com-p.123">
      <name full="Freddy Garcia" role="nrol: full">
        Freddy Garcia
      </name>
      <name part="nprt: given">
        Freddy
      </name>
      <name part="nprt: family">
        Garcia
      </name>
    </player-metadata>
    <player-stats>
      <player-stats-baseball>
        <stats-baseball-pitching earned-runs="4.81" wins="1" losses="3"
          winning-percentage=".250"/>
      </player-stats-baseball>
    </player-stats>
  </player>
</team>
```

### 13.3.5.4 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>
```

### 13.3.5.5 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" />
```

### 13.3.5.6 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>
```

### 13.3.5.7 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 Michel.
  <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

### 13.3.5.8 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>
```

### 13.3.5.9 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 19 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="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/
    NewsML-G2_2.2-spec-NewsItem-Power.xsd"
  xmlns:xts="http://www.xml.team.com" standard="NewsML-G2" standardversion="2.2"
  conformance="power" xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_1.xml" />
  <catalogRef
    href="http://www.xml.team.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="ni: nat: text" />
    <provider qcode="web: www.xml.team.com" />
    <versionCreated>2009-02-02T16:00:00Z</versionCreated>
    <pubStatus qcode="stat: usable" />
    <fileName>xt.5932656-pitcher-preview.xml</fileName>
  </itemMeta>
  <contentMeta>
    <contentCreated>2009-02-02T12:17:00-05:00</contentCreated>
    <located qcode="ci: ty: Philadelphia">
      <broadener qcode="reg: Pennsylvani a" />
      <broadener qcode="cntry: USA" />
    </located>
    <creator qcode="web: sportsnetwork.com">
      <name>The Sports Network</name>
    </creator>
    <altId type="idtype: tsn-id" id="sportsnetwork.com-5932656" />
    <altId type="idtype: revision-id"
      id="l.mlb.com-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
    <genre type="spcpnat: spfixt" qcode="spfixt: pitcher-preview">
      <name xml:lang="en-US">Game Pitcher Preview</name>
      <broadener 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>
    <subject qcode="subj: 15007000">
      <name xml:lang="en-US">Baseball</name>
      <broadener qcode="subj: 15000000" />
    </subject>
    <subject qcode="league: l.mlb.com">
      <name xml:lang="en-US">Major League Baseball</name>
      <broadener qcode="subj: 15007000" />
    </subject>
```



```

    <sameAs qcode="subj: 15007001" />
  </subject>
  <subject type="spcpnat: conf" qcode="conf: I. ml b. com-c. national ">
    <name xml: lang="en-US">National </name>
    <broadier qcode="league: I. ml b. com" />
  </subject>
  <subject type="cpnat: event" qcode="event: I. ml b. com-2007-e. 19358" />
  <subject type="spcpnat: team" qcode="team: I. ml b. com-t. 19">
    <name>Phi I adel phi a Phi I I i es</name>
  </subject>
  <subject type="spcpnat: team" qcode="team: I. ml b. com-t. 26">
    <name>Ari zona Di amondbacks</name>
  </subject>
  <subject type="cpnat: person" qcode="person: I. ml b. com-p. 456">
    <name rol e="nrol: full">Doug Davi s</name>
    <sameAs qcode="fssID: 45679" />
  </subject>
  <subject type="cpnat: person" qcode="person: I. ml b. com-p. 123">
    <name rol e="nrol: full">Freddy Garci a</name>
    <sameAs qcode="fssID: 45680" />
  </subject>
  <headline>Pi tcher Preview: Ari zona Di amondbacks (29-23) at Phi I adel phi a
    Phi I I i es (26-24), 7: 05 p. m. </headline>
  <slugline separator="-">AAV! PREVI EW-ARI -PHI </slugline>
</contentMeta>
<contentSet>
  <inlineXML contenttype="appl i cati on/sportsml +xml ">
    <sports-content xml ns="http: //i ptc. org/std/SportsML/2008-04-01/"
      xsi: schemaLocati on="http: //i ptc. org/std/SportsML/2008-04-01/
        sportsml -G2. xsd">
      <sports-event>
        <event-metadata date-coverage-type="event"
          event-key="event: I. ml b. com-2009-e. 19358"
          date-coverage-val ue="I. ml b. com-2009-e. 19358"
          event-status="pre-event" start-date-ti me="2009-02-06T19: 05: 00-05: 00">
        </event-metadata>
        <team>
          <team-metadata team-key="team: I. ml b. com-t. 26"
            al i gnment="away">
            <name full ="Ari zona Di amondbacks" rol e="nrol: full ">
              Ari zona Di amondbacks</name>
            </team-metadata>
            <player>
              <player-metadata posi ti on-event="1"
                pl ayer-key="pl ayer: I. ml b. com-p. 456">
                <name full ="Doug Davi s" rol e="nrol: full ">Doug Davi s</name>
                <name part="nprr: gi ven">Doug</name>
                <name part="nprr: fami ly">Davi s</name>
              </player-metadata>
              <player-stats>
                <player-stats-basebal l>
                  <stats-basebal l -pi tchi ng earned-runs="3. 57"
                    wi ns="2" l osses="6" wi nni ng-percentage=" . 250" />
                </player-stats-basebal l>
              </player-stats>
            </player>
          </team>
          <team>
            <team-metadata team-key="team: I. ml b. com-t. 19"
              al i gnment="home">
              <name full ="Phi I adel phi a Phi I I i es" rol e="nrol: full ">
                Phi I adel phi a Phi I I i es</name>
              </team-metadata>
              <player>
                <player-metadata posi ti on-event="1"
                  pl ayer-key="pl ayer: I. ml b. com-p. 123">
                  <name full ="Freddy Garci a" rol e="nrol: full ">
                    Freddy Garci a</name>
                  <name part="nprr: gi ven">Freddy</name>
                  <name part="nprr: fami ly">Garci a</name>
                </player-metadata>
                <player-stats>
                  <player-stats-basebal l>
                    <stats-basebal l -pi tchi ng earned-runs="4. 81"
                      wi ns="1" l osses="3" wi nni ng-percentage=" . 250" />
                  </player-stats-basebal l>
                </player-stats>
              </player>
            </team>

```



```
        </sports-event>
      </sports-content>
    </inlineXML>
  </contentSet>
</newsItem>
```

### 13.3.6 SportsML-G2 Packages

To convey a package of SportsML-G2 items, simply use the 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 (NITF<sup>13</sup>) document, conveyed in a NewsML-G2 News Item, as shown in **LISTING 20**.

#### LISTING 20 Sports story in NewsML-G2/NITF

```
<?xml version="1.0" encoding="UTF-8"?>
<newsItem
  guid="tag:xml.team.com,2009:xt.5932656-preview"
  version="1"
  xmlns="http://iptc.org/std/nar/2006-10-01/"
  xsi:schemaLocation="http://www.w3.org/2001/XMLSchema-instance
http://iptc.org/std/nar/2006-10-01/NewsML-G2.2-spec-NewsItem-Power.xsd"
  xmlns:xts="http://www.xml.team.com/xts"
  standard="NewsML-G2"
  standardversion="2.2"
  conformance="power"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_1.xml" />
  <catalogRef
    href="http://www.xml.team.com/specification/xts-SportsCodesCatalog_1.xml" />
  <itemMeta>
    <itemClass qcode="ni:news-text" />
    <provider qcode="web:www.xml.team.com" />
    <versionCreated>2009-02-02T11:45:00-04:00</versionCreated>
    <pubStatus qcode="stat:usable" />
    <fileName>xt.5932656-preview.xml</fileName>
  </itemMeta>
  <contentMeta>
    <contentCreated>2009-02-02T11:15:00-05:00</contentCreated>
    <located qcode="ci:ty:Philadelphia">
      <broaden qcode="reg:Pennsylvania" />
      <broaden qcode="cntry:USA" />
    </located>
    <creator qcode="web:sportsnetwork.com">
      <name>The Sports Network</name>
    </creator>
    <altId type="idtype:tsn-id" id="sportsnetwork.com-5932656" />
    <altId type="idtype:revision-id" id="1.mlb.com-2007-e.19358-pre-event-coverage-sportsnetwork.com" />
    <genre type="spcpnat:spfxt" qcode="spfxt:pre-event-coverage">
      <name xml:lang="en-US">Game Preview</name>
      <broaden qcode="dclass:event-summary" />
    </genre>
    <genre type="spcpnat:dclass" qcode="dclass:event-summary" />
    <genre type="tsn-genre:tsn-structure" qcode="tsn-structure:mlbpreviewxml" />
    <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>
      <broaden qcode="subj:15000000" />
    </subject>
    <subject qcode="league:1.mlb.com">
      <name xml:lang="en-US">Major League Baseball</name>
      <broaden 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>
```

<sup>13</sup> The IPTC's XML standard for marking up text.





```
<broader qcode="league: I. ml b. com" />
</subject>
<subject type="cpnat: event" qcode="event: I. ml b. com-2007-e. 19358" />
<subject type="spcpnat: team" qcode="team: I. ml b. com-t. 19">
  <name>Philadelphia Phillies</name>
</subject>
<subject type="spcpnat: team" qcode="team: I. ml b. com-t. 26">
  <name>Arizona Diamondbacks</name>
</subject>
<subject type="cpnat: person" qcode="person: I. ml b. com-p. 456">
  <name>Doug Davis</name>
  <sameAs qcode="fssID: 45679" />
</subject>
<subject type="cpnat: person" qcode="person: I. ml b. com-p. 123">
  <name>Freddie Garcia</name>
  <sameAs qcode="fssID: 45680" />
</subject>
<headline>Preview: Arizona Diamondbacks (29-23) at Philadelphia Phillies
(26-24), 7:05 p.m.</headline>
<slugline separator="-">AAV! PREVIEW-ARI-PHI</slugline>
<description role="drol: summary">A pair of teams coming off big weekend sweeps
will square off tonight at Citizens Bank Park, where the Philadelphia Phillies
welcome the Arizona Diamondbacks for the start of a three-game
series.</description>
</contentMeta>
<contentSet>
  <inlineXML contenttype="application/nitf+xml">
    <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/"
      xsi:schemaLocation="http://iptc.org/std/NITF/2006-10-18/
        nitf-3-4.xsd">
      <body>
        <body.head>
          <headline>
            <hl1>Arizona Diamondbacks (29-23) at Philadelphia Phillies (26-24),
              7:05p.m.</hl1>
          </headline>
          <byline>
            <byttl>Sports Network</byttl>
          </byline>
          <abstract>
            <p>A pair of teams coming off big weekend sweeps will square off
              tonight at Citizens Bank Park, where the Philadelphia Phillies
              welcome the Arizona Diamondbacks for the start of a three-game
              series.</p>
          </abstract>
        </body.head>
        <body.content>
          <p>(Sports Network) - A pair of teams coming off big weekend sweeps
            will square off tonight at Citizens Bank Park, where the Philadelphia
            Phillies welcome the Arizona Diamondbacks for the start of a three-game
            series.</p>
          <p>The Phillies picked up some momentum with a three-game sweep of the
            Atlanta Braves, while the Diamondbacks captured all four of their games
            against the Houston Astros.</p>
          <p>On Sunday, Livan Hernandez threw his first complete game in almost
            two years to lead the Diamondbacks to an 8-4 win over the Astros to
            complete the sweep.</p>
          <p>Arizona sends Doug Davis to the hill tonight, and the lefty is 0-4
            over his last five starts. That includes losses in each of his last
            three outings.</p>
          <p>The Phillies, meanwhile, notched their first sweep of the year that
            culminated with Sunday's 13-6 pounding of the Braves. Ryan Howard
            flashed some of the power that won him the NL MVP last season, jacking
            a pair of two-run homers in the win.</p>
          <p>Right-hander Freddie Garcia takes the hill for the Phillies and is 1-
            3 with a 4.81 earned run average on the year. Since winning his first
            game in a Philadelphia uniform back on April 22, Garcia has gone 0-2
            over his last six starts.</p>
        </body.content>
      </body>
    </nitf>
  </inlineXML>
</contentSet>
</newsItem>
```





### 13.3.6.1 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 G2 Package Item.

### 13.3.6.2 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 G2 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
    NAR_1.3-spec-PackageItem-Power.xsd "
  xmlns:xts="www.xmlteam.com"
  conformance="power"
  standard="NewsML-G2" ← standard and
  standardversion="2.2" standard version
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
```

### 13.3.6.3 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.

### 13.3.6.4 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 attribute 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 25**

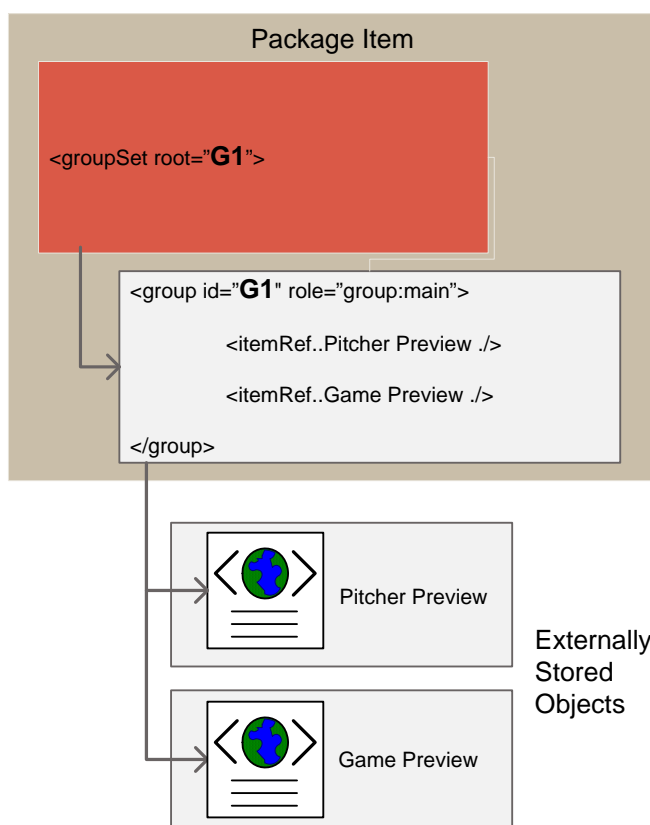


Figure 25: Diagram of Package Group structure

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:spfi xt" qcode="spfi xt: pitcher-preview">
        <name xml:lang="en-US">Game Pitcher Preview</name>
        <broadier qcode="dcl ass: event-summary" />
      </genre>
      <genre type="spcpnat:dcl ass" qcode="dcl ass: event-summary" />
      <genre type="xts-genre:tsn-fi xture" qcode="tsn-fi xture: ml bprevi ewxml " />
      <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/nift+xml">
      <genre type="spcpnat:spfi xt" qcode="spfi xt: pre-event-coverage">
        <name xml:lang="en-US">Game Preview</name>
        <broadier qcode="dcl ass: event-summary" />
      </genre>
      <genre type="spcpnat:dcl ass" qcode="dcl ass: event-summary" />
      <genre type="xts-genre:tsn-fi xture" qcode="tsn-fi xture: ml bprevi ewxml " />
      <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 21 SportsML-G2 Package

```
<?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/
    NAR_1.3-spec-PackageItem-Power.xsd"
  xmlns:xts="http://www.xmlteam.com"
  conformance="power"
  standard="NewsML-G2"
  standardversion="2.2"
  xml:lang="en-US">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_6.xml" />
  <catalogRef
    href="http://www.xmlteam.com/specification/xts-SportsCodesCatalog_1.xml" />
  <iItemMeta>
    <iItemClass qcode="ni nat:composi te" />
    <provider qcode="web:xml team.com" />
    <versionCreated>2009-02-02T15:00:09Z</versionCreated>
  </iItemMeta>
  <contentMeta>
    <contentCreated>2009-02-02T11:45:00-05:00</contentCreated>
    <located qcode="ci ty:Phi l adel phi a">
      <broader qcode="reg:Pennsyl vani a" />
      <broader qcode="cn try:USA" />
    </located>
    <creator qcode="web:sportsnetwork.com">
      <name>The Sports Network</name>
    </creator>
    <altId type="i dtype:tsn-id" id="sportsnetwork.com-5932656" />
    <altId type="i dtype:revi si on-id"
      id="l.ml b.com-2009-e.19358-pre-event-coverage-sportsnetwork.com" />
    <genre type="spcpnat:dcl ass" qcode="dcl ass: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="l 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="spcpnat:conf" qcode="conf:l.ml b.com-c.nati onal">
      <name xml:lang="en-US">Nati onal</name>
      <broader qcode="l league:l.ml b.com" />
    </subject>
    <subject type="cpnat:event" qcode="event:l.ml b.com-2007-e.19358" />
    <subject type="spcpnat:team" qcode="team:l.ml b.com-t.19">
      <name>Phi l adel phi a Phi l l i es</name>
    </subject>
    <subject type="spcpnat:team" qcode="team:l.ml b.com-t.26">
      <name>Ari zona Di amondbacks</name>
    </subject>
  </contentMeta>
  <groupSet root="G1">
    <group id="G1" role="group:mai n">
      <iItemRef href="xt.5932656-pi tcher-previ ew"
        contenttype="appl i cati on/sportsml+xml">
        <genre type="spcpnat:spfi xt" qcode="spfi xt:pi tcher-previ ew">
          <name xml:lang="en-US">Game Pi tcher Prev i ew</name>
          <broader qcode="dcl ass:event-summary" />
        </genre>
        <genre type="spcpnat:dcl ass" qcode="dcl ass:event-summary" />
        <genre type="xts-genre:tsn-fi xt ure" qcode="tsn-fi xt ure:ml bprevi ewxml" />
        <headline>Pi tcher Prev i ew: Ari zona Di amondbacks (29-23) at Phi l adel phi a
          Phi l l i es (26-24), 7:05 p.m.
        </headline>
        <descri pti on></descri pti on>
      </iItemRef>
      <iItemRef href="xt.5932656-previ ew" contenttype="appl i cati on/nif+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:mlbpreviewxml" />
<headline>Preview: Arizona Diamondbacks (29-23) at Philadelphia
  Phillies (26-24), 7:05 p.m.
</headline>
<description role="drol:summary">A pair of teams coming off big weekend
  sweeps will square off tonight at Citizens Bank Park, where the
  Philadelphia Phillies welcome the Arizona Diamondbacks for the start of a
  three-game series.
</description>
</itemRef>
</group>
</groupSet>
</packageItem>

```



This page intentionally blank

## 14 Exchanging News: News Messages

### 14.1 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 G2 Items by any kind of digital transmission system.

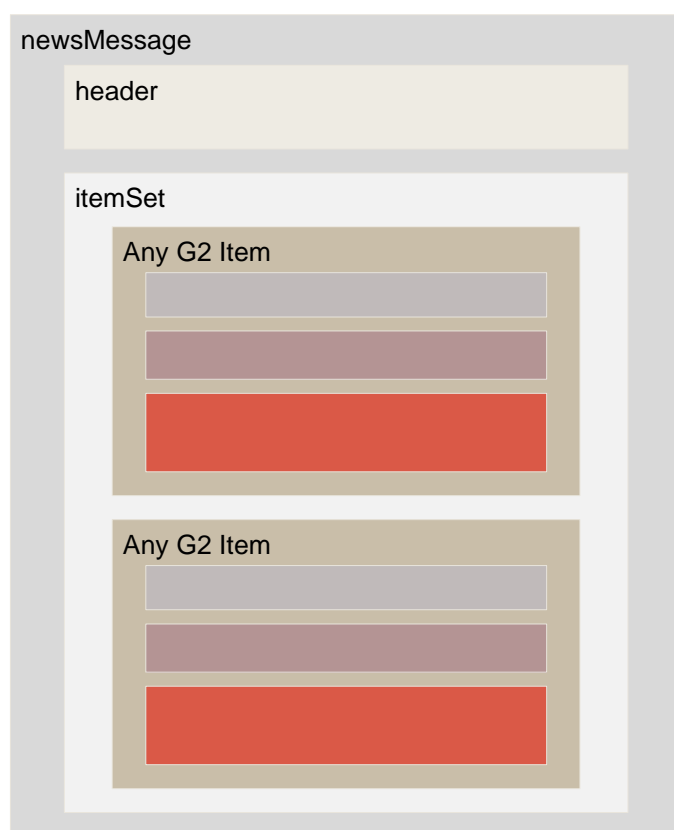
The use of <newsMessage> is entirely optional: items may also be exchanged using SOAP, Atom Publishing Protocol (APP) or other suitable content syndication method.

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.

### 14.2 Structure

A diagram of the structure of a News Message and its payload of G2 Items is shown in **Figure 26** below.



**Figure 26: 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.



## 14.2.1 News Message Header <header>

The Header contains a set of properties useful for managing the transmission and reception of news.

### 14.2.1.1 Sent timestamp <sent>

Mandatory, non-repeatable. The transmission timestamp is in XML date-time 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>
```

### 14.2.1.2 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>
```

### 14.2.1.3 Transmission ID <transmitId>

An optional, non-repeatable string identifying the news message. The ID must be unique for every message on a given calendar day, except that if a message is re-transmitted, it may keep the same ID. The structure of the string is not specified by the IPTC.

```
<transmi tI d>87e3841c-430e-4da5-bed6-28cae4e8fe3b</transmi tI D>
```

### 14.2.1.4 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. Note that this is not the same as the <urgency> of an individual Item within the news message, although they may be correlated. Priority would be used by a transmission system to determine the order in which a message is transmitted, relative to others in a queue. Urgency is an editorial judgement expressing the relative news value of an item.

```
<pri ori ty>4</pri ori ty>
```

### 14.2.1.5 Message Origin <origin>

An optional, non-repeatable string whose structure is not specified by the IPTC. It could denote the name of a channel, system or service, for example.

```
<ori gi n>MMS_3</ori gi n>
```

### 14.2.1.6 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:

```
<desti nati on>USA</desti nati on>  
<desti nati on>UK, France</desti nati on>
```

### 14.2.1.7 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 or service – a virtual <channel> delivered within a physical stream identified by <destination>

```
<channel >TVS</channel >  
<channel >TTT</channel >  
<channel >WWW</channel >
```

The diagram shown below in **Figure 27** illustrates the application of Destination and Channel in a syndication workflow.



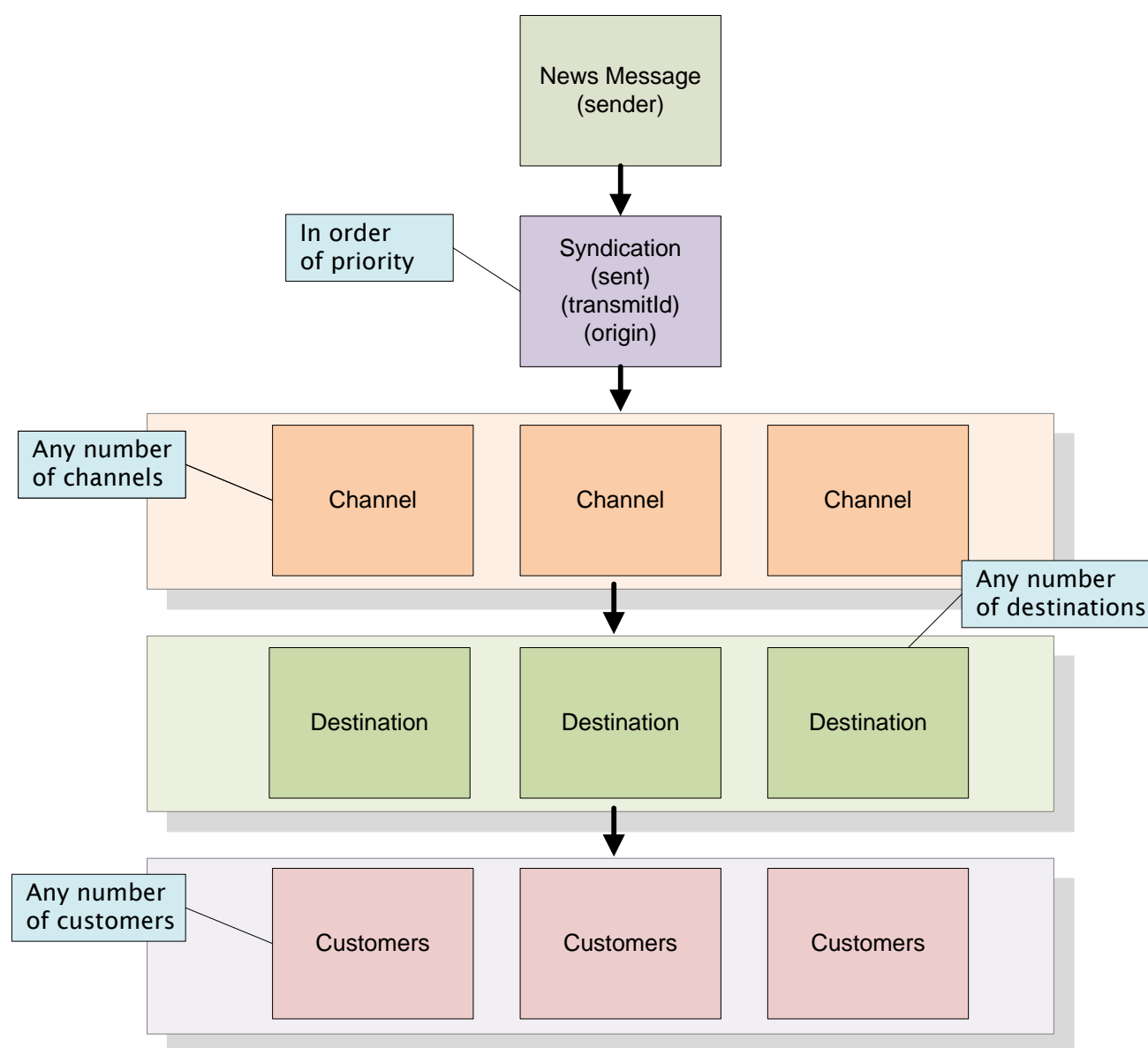


Figure 27: Syndication workflow using Destination and Channel

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.

#### 14.2.1.8 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 optional time zone. The property may be refined using the optional @role, which is a string value.<sup>14</sup>

```

<timestamp role="received">2009-02-01T11:17:00.000Z</timestamp>
<timestamp role="transmitted">2009-02-01T11:17:00.100Z</timestamp>

```

#### 14.2.2 News Message payload <itemSet>

Each <newsMessage> MUST have one <itemSet> component, which wraps the G2 Items that are to be transmitted.

<sup>14</sup> None of the News Message properties can use QCode values, since there is no catalog or scheme URI information present within the scope of the <header> element.



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.

#### LISTING 22 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/
    NAR_1.3-spec-NewsMessage-Power.xsd">
  <header>
    <sent>2009-02-01T11:17:00.100Z</sent>
    <sender>thomsonreuters.com</sender>
    <transmissionId>tag:reuters.com,2009:newsml_OVE488500-PKG</transmissionId>
    <priority>4</priority>
    <origin>MMS_3</origin>
    <destination>UKI</destination>
    <channel>TVS</channel>
    <channel>TTT</channel>
    <channel>WWW</channel>
    <timestamp role="received">2009-02-01T11:17:00.000Z</timestamp>
    <timestamp role="transmitted">2009-02-01T11:17:00.100Z</timestamp>
  </header>
  <itemSet>
    <packageItem>
      <itemRef residref="N1" />
      <itemRef residref="N2" />
      <itemRef residref="N3" />
      <itemRef residref="N4" />
    </packageItem>
    <newsItem guid="N1">.....</newsItem>
    <newsItem guid="N2">.....</newsItem>
    <newsItem guid="N3">.....</newsItem>
    <newsItem guid="N4">.....</newsItem>
  </itemSet>
</newsMessage>
```



## 15 Migrating IPTC 7901 to G2

### 15.1 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 or G2 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 G2.

When providers convert to 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 G2 property to which the 7901 timestamp should be mapped.

Customers migrating to G2 and converting IPTC 7901 fields to G2 properties would need to consult their providers' documentation, or make inquiries about the mapping of IPTC 7901 fields in the provider's systems.

#### 15.1.1 Message Header

Field Name	Example	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 G2 property would be the &lt;destination&gt; property of News Message:</p> <pre>&lt;destination&gt;FRS&lt;/destination&gt;</pre> <p>An alternative property would be &lt;channel&gt;</p> <pre>&lt;channel&gt;FRS&lt;/channel&gt;</pre> <p>Within a News Item, there is a &lt;service&gt; 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 &lt;transmitId&gt; of News Message</p>
Priority of Story	3	<p>IPTC 7901 specifies a single digit of value 1 to 6 inclusive. G2 has two equivalents:</p> <pre>&lt;urgency&gt; in the Administrative Metadata group of &lt;contentMeta&gt;</pre> <pre>&lt;urgency&gt;3&lt;/urgency&gt;</pre> <pre>&lt;priority&gt; in the &lt;header&gt; of News Message.</pre> <pre>&lt;priority&gt;3&lt;/priority&gt;</pre> <p>If the 7901 field is mapped from an editorial system and is under the control of journalists, then &lt;urgency&gt; 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	In IPTC7901, this could be up to three characters. A provider converting its



Field Name	Example	G2 Property and Example
		<p>categories for 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 &lt;subject&gt; in Content Metadata</p> <pre>&lt;subject type="cpnat: abstract" qcode="cat: P0L"&gt;   &lt;name&gt;Pol i t i c s&lt;/name&gt; &lt;/subj ect&gt;</pre>
Word Count	195	<p>The maximum value allowed in IPTC 7901 is 9999. Word Count is one of the properties of <b>News Content Characteristics</b> in G2 and for the plain text of IPTC 7901 is an attribute of the Inline Data component &lt;inlineData&gt;</p> <pre>&lt;i n l i n e D a t a wordcount=" 195" &gt; Text text text &lt;/i n l i n e D a t a&gt;</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 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 G2 property.</p>
Keyword/Catch-Line		<p>Customers converting an IPTC 5901 service into G2 and who are unsure about which of 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 G2 &lt;slugline&gt; property in the Descriptive Metadata group of &lt;contentMeta&gt;.</p> <pre>&lt;sl u g l i n e separator=" - " &gt;US-P0LI TI CS/BUSH&lt;/sl u g l i n e&gt;</pre>

### 15.1.2 Message Text

If the message text is to be conveyed as plain text, then the <inlineData> wrapper would be used. If the migration to G2 also involves migration of the text to some mark-up such as NITF or xhtml, this would be conveyed in the <inlineXML> wrapper.

### 15.1.3 Post-Text Information

The IPTC7901 post-text fields all relate to date-time. As previously recommended, providers converting to G2 should examine the origin of the information being used to fill the IPTC 7901 day and time field, and use the appropriate 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 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 optional 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 time and time zone optional.

| <contentCreated>2009-02-09</contentCreated>

Field Name	Example	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 G2 (none needed)
Month of Transmission	Feb	Three-alpha character field. Optional in IPTC 7901. No separate equivalent in G2 (none needed)
Year of Transmission	09	Two-digit field. Optional in IPTC 7901. No separate equivalent in G2 (none needed)



This page intentionally blank

## 16 News Industry Text Format (NITF)

### 16.1 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; G2 makes more use of generic structures that can be adapted for different uses. Therefore there may not be a direct equivalent in 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 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 G2.



## 16.2 Overview of NITF and 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.

### 16.2.1 Root element <nitf>

The root element may uniquely identify the document using @uno, which is equivalent to G2 @guid property

### 16.2.2 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 G2 <itemMeta> component, with exceptions noted in the following table:

NITF element	G2 equivalent	Comment
<title>	<title>	The properties are equivalent in NITF and G2
<meta>	<itemClass>	<itemClass> uses an IPTC NewsCode to identify the type of content being conveyed.
<tobject>	<subject>	This is part of G2 <contentMeta>
<iim>	None	Deprecated in NITF, no G2 equivalent
<docdata>	Various	A container for detailed metadata for the document. contains child elements as tabled in <a href="#">16.2.2.1</a> below
<pubdata>	None	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

#### 16.2.2.1 <docdata> child elements

These convey detailed management information about the document.

NITF element	G2 equivalent	Comment
<correction>	<signal>	Indicates that the item is a correction, identifies the previously published document and any correction instruction. 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 G2 equivalent of <subject> may use a @role to indicate that the subject is a geographic place. If used, the IPTC Concept Type NewsCode would be either "geoArea" or "poi" (Point of Interest)



NITF element	G2 equivalent	Comment
<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 G2 News Message <header>.
< urgency>	< urgency>	Editorial urgency of the document. NITF uses values 1-8, G2 uses 1-9. The value may also be linked to Priority <priority> in a G2 News Message.
< fixture>	< instanceOf>	Regularly occurring and/or frequently updated content, for example "Noon Weather Report" or "Closing Stock Market Prices". In 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 G2 property would be <versionCreated>, but depending on the provider's intention, <firstCreated>, <contentCreated> or <contentModified> could be used.
< date.release>	Various	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 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 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 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 G2
< du-key>	None	NITF uses this property for maintaining a network of relationships to other versions of the same article. G2 uses GUID, versioning and, if necessary, Links, to maintain this information
<doc.copyright >	< rightsInfo>	In 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>	G2 has no specific equivalent container for keywords.



NITF element	G2 equivalent	Comment
		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 G2 equivalent &lt;inlineRef&gt; enables a G2 Concept to be referenced by @idrefs, and contains a QCode identifying the Concept in a controlled vocabulary. Locally, any mark-up element (e..g &lt;span&gt; in XHTML capable of using @id may be linked via &lt;inlineRef&gt; to the concept.</p>

### 16.3 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 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 G2 equivalents.

NITF element	G2 equivalent	Comment
<headline>	<headline >	G2 <headline> is contained in <contentMeta> and may appear any number of times. It can be refined by using @role to express the NITF notions of main headline <hl1> and subhead <hl2>
<note >	Various	<p>An advisory for editors may be equivalent to G2 &lt;edNote&gt;in &lt;itemMeta&gt;, but NITF &lt;note&gt; has a number of uses, including expressing copyright, and may also be publishable.</p> <p>Provider's migrating from NITF to G2 will need to identify the intention of their use of &lt;note&gt; and choose from several G2 properties that express this intention, for example &lt;rightsInfo&gt; (a property of the G2 root element), or &lt;description&gt; (part of &lt;contentMeta&gt;) with a @role defining its purpose.</p> <p>Customers converting an NITF service to G2 will need to consult with their provider, either directly or via documentation.</p>
<rights >	< rightsInfo>	See <a href="#">16.2.2.1</a> above
<byline >	<by>	If using the every available feature of NITF <byline>, G2 <by> should be implemented at PCL. It is part of <contentMeta>
<distributor >	<provider >	<provider> is a mandatory property of G2 <itemMeta>
<dateline >	< dateline>	G2 <dateline> is part of <contentMeta>
< abstract>	< description>	G2 <description> is part of <contentMeta> and may be refined by @role. The IPTC Description Role NewsCode value would be "summary"
<series >	<memberOf>	See <a href="#">16.2.2.1</a> above



## 16.4 Body Content <body.content>

G2 has no equivalent mark-up for text, and NITF is a valid choice for marking up text to be conveyed by G2. An NITF mark-up would be conveyed in G2 <inlineXML>.

### LISTING 23 An NITF marked-up article conveyed in G2 <inlineXML>

```
<contentSet>
  <inlineXML contenttype="application/nitf+xml">
    <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/"
      xsi:schemaLocation="http://iptc.org/std/NITF/2006-10-18/
http://iptc.org/std/NITF/3.4/specification/schema/nitf-3-4.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>
```

## 16.5 End of Body <body.end>

There is no formal equivalent to the <body.end> container in 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 by G2, or expressed in another mark-up language.

The G2 equivalent of <tagline> would be <by> with a @role if necessary to refine its use. <bibliography> may be expressed using G2 <description> with a @role. Their placement relative to the parent text is a processing issue – G2 is format-agnostic.



This page intentionally blank

## 17 Mapping Embedded Photo Metadata to G2

### 17.1 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 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 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](http://www.metadataworkinggroup.org).

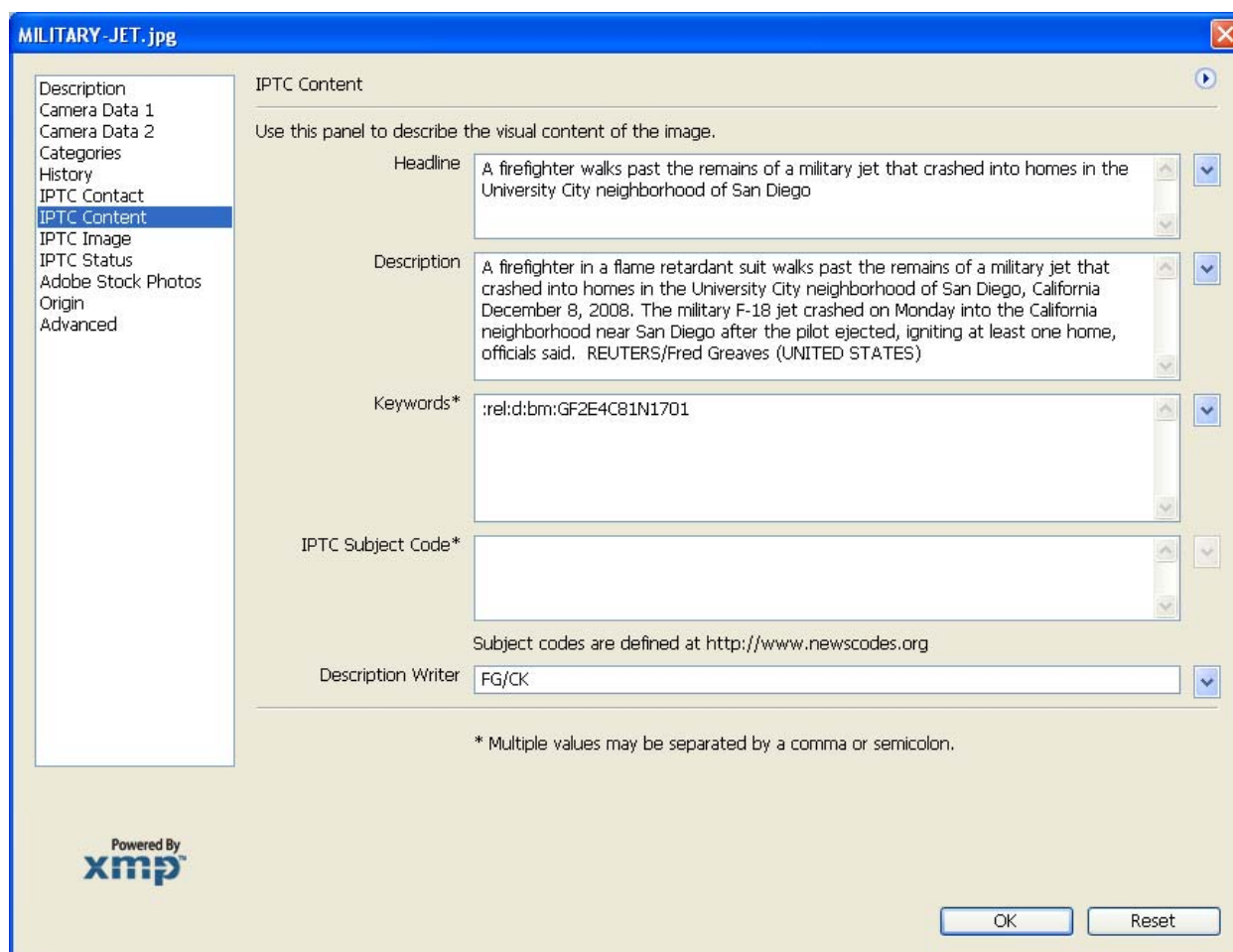


Figure 28: One of the IPTC Core metadata panels from an image opened in Adobe Photoshop CS2.

### 17.2 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 [www.iptc.org/iptc4xmp](http://www.iptc.org/iptc4xmp).

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 ([http://www.iptc.org/std/lptc4xmpCore/1.0/documentation/lptc4xmpCore\\_1.0-doc-CpanelsUserGuide\\_13.pdf](http://www.iptc.org/std/lptc4xmpCore/1.0/documentation/lptc4xmpCore_1.0-doc-CpanelsUserGuide_13.pdf)) tabulates the field labels used by IIM, IPTC Core, and the equivalent fields of software packages such as Photoshop, iViewMedia, Thumbs, and Irfanview.

### 17.3 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.

### 17.4 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.

### 17.5 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 G2 Standards:

- ❖ metadata carried in a 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 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 G2 Standards. Some are already using 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.





## 17.6 IIM resources

The IPTC web site has an IIM Home page ([www.iptc.org/iim](http://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](http://www.controlledvocabulary.com).

More about the use of IIM for photos can be found in the “Photo Metadata” section of the IPTC website.

## 17.7 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 G2 metadata properties. It is NOT intended to describe a mapping of the complete IIM envelope to G2.

The mapping of IPTC Core Schema properties to G2 is documented in the IPTC Photo Metadata Specification 2009, which can be found at: <http://www.iptc.org/std/photometadata/specification/>

## 17.8 IIM to 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 [15.4.3](#)

DataSet	IIM Name	IPTC Core Name (XMP)	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/facet	
2:90	City	City	contentMeta/subject	May be used with <assert> (see below)
2:92	Sub-location	Location		
2:95	Province/State	State		
2:100	Country/Primary Location Code	Country Code		
2:101	Country/Primary Location Name	Country		
2:103	Original Transmission Reference	Transmission Reference	itemMeta/altId	If the reference is a Job ID use itemMeta/memberOf
2:105	Headline	Headline	contentMeta/headline	
2:110	Credit	Credit	contentMeta/creditline	
2:115	Source	Source	rightsInfo/copyrightHolder	
2:116	Copyright Notice	Copyright Notice	rightsInfo/copyrightNotice	

DataSet	IIM Name	IPTC Core Name (XMP)	G2 Property	Note
2:120	Caption/Abstract	Description	contentMeta/description	
2:122	Writer/Editor	Caption Writer	contentMeta/contributor	A @role should be added to contributor

## 17.9 IIM to 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 G2, and their values, are tabulated, and in the succeeding paragraphs, the mapping discussed in more detail.

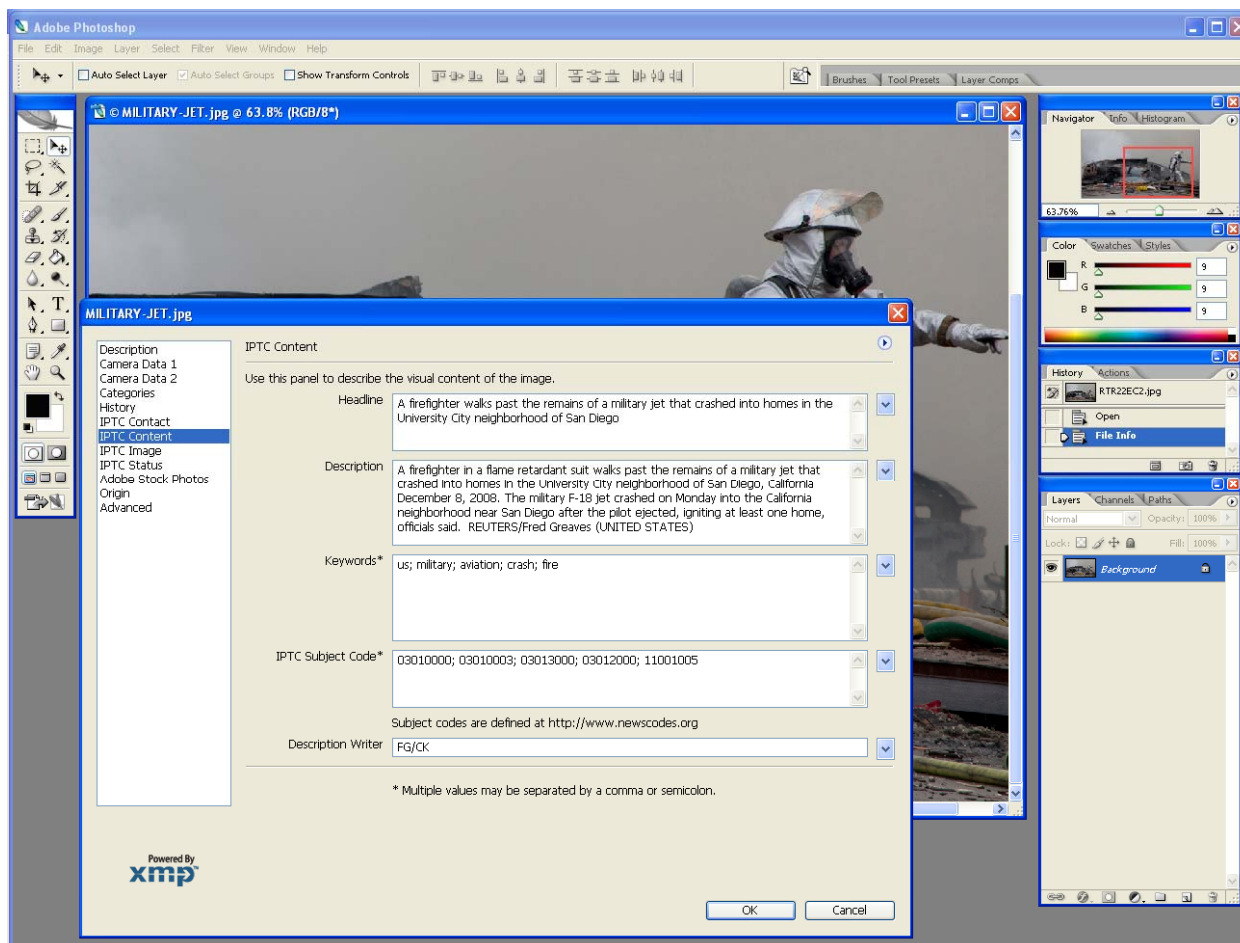


Figure 29: Opening File Info panel of the image (photo and metadata ©Thomson Reuters)

DataSet	IIM Name	Value	Ref
2:05	Object Name	USA-CRASH/MILITARY	17.9.1
2:10	Urgency	4	17.9.2
2:15	Category	A	17.9.3
2:20	Supplemental Category	DEF DIS	17.9.3
2:25	Keywords	us	17.9.3
2:25	Keywords	military	
2:25	Keywords	aviation	
2:25	Keywords	crash	
2:25	Keywords	fire	
2:40	Special Instruction	NO ARCHIVE OR WIRELESS USE	17.9.5
2:55	Date Created	20081208	17.9.6



DataSet	IIM Name	Value	Ref
2:60	Time Created	133000-0800	<a href="#">17.9.6</a>
2:80	By-Line	John Smith	<a href="#">17.9.7</a>
2:85	By-line Title	Staff Photographer	<a href="#">17.9.7</a>
2:90	City	San Diego	<a href="#">17.9.8</a>
2:92	Sub-location	University City	<a href="#">17.9.8</a>
2:95	Province/State	CA	<a href="#">17.9.8</a>
2:100	Country/Primary Location Code	USA	<a href="#">17.9.8</a>
2:101	Country/Primary Location Name	United States	<a href="#">17.9.8</a>
2:103	Original Transmission Reference	SAD02	<a href="#">17.9.9</a>
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	<a href="#">17.9.10</a>
2:110	Credit	Acme/John Smith	<a href="#">17.9.7</a>
2:115	Source	Acme News LLC	<a href="#">17.9.7</a>
2:116	Copyright Notice	© Copyright Acme News LLC. Click For Restrictions - <a href="http://www.acmenews.com/legal">http://www.acmenews.com/legal</a>	<a href="#">17.9.11</a>
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.	<a href="#">17.9.12</a>
2:122	Writer/Editor	CK	<a href="#">17.9.13</a>

### 17.9.1 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="ni nat: pi ctu re" />
  <provider literal="reuters.com" />
  <versionCreated>
    2008-12-09T02: 20: 00Z
  </versionCreated>
  <title>USA-CRASH/MI LI TARY</ti tle>  ← Title
</i temMeta>
```

Some providers may additionally map this DataSet to the G2 <slugline> property in <contentMeta>:

```
<slugline>USA-CRASH/MI LI TARY</slugline>
```

### 17.9.2 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 G2 property, which is a child of <contentMeta>:

```
<urgency>4</urgency>
```



### 17.9.3 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: DI S" />
```

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.

### 17.9.4 Keywords

A specific Keyword property was added to G2 in NewsML-G2 version 2.4. 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. 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 G2 <keyword>, the IPTC suggests the following rules when configuring a mapping of Keywords metadata to G2:

- ❖ Assess if any existing G2 properties align to the use of this 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.

### 17.9.5 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>
...
```



```
... </title>
<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>
```

### 17.9.6 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 G2 property accepts a Truncated DateTime value (i.e. the value may be truncated if parts of the Date-Time are not available).

```
<contentMeta>
<contentCreated>2008-02-08T13:30:00-08:00</contentCreated>
...
</contentMeta>
```

### 17.9.7 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 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 <facet> 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>
<facet qcode="fcode:jobtitle">
<name>Staff Photographer</name>
</facet>
</creator>
...
</contentMeta>
```

Credit should be mapped to the <creditline> child property of <contentMeta>. There is a <provider> property of <itemMeta>, but the Credit does necessarily reflect the provider. Many picture providers use IIM Credit to display the name of the person, organisation, or both, who should be credited when the picture is used. In this context, <creditline> is appropriate because it is a natural-language label that is intended to be displayed.





```
<creditline>Acme/John 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.

The G2 structure 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>
```

### 17.9.8 City, Province/State and Country

As discussed in [6 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. G2 has a <located> property to express this.
- ❖ The location shown in the picture – in G2 this is one of the <subject> properties 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, facet, 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: geoAreaSubl ocation" literal="int001">
  <name>Uni versi ty Ci ty</name>
  <broader type="cpnatexp: geoAreaCi ty" literal="int002" />
</subject>
<subject type="cpnatexp: geoAreaCi ty" literal="int002">
  <name>San Di ego</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>Uni ted 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: geoAreaSubl ocati on" l i t e r a l ="i n t 0 0 1" />  ← @ l i t e r a l
</contentMeta>
<assert l i t e r a l ="i n t 0 0 1">  ← @ l i t e r a l
  <type qcode="cpnatexp: geoAreaSubl ocati on" />
  <name>Uni versi ty Ci ty</name>
  <broaden l i t e r a l ="i n t 0 0 2" type="cpnatexp: geoAreaCi ty">
    <name>San Di ego</name>
  </broaden>
  <broaden l i t e r a l ="i n t 0 0 3" type="cpnatexp: geoAreaProvState">
    <name>Cal i forni a</name>
  </broaden>
  <broaden l i t e r a l ="USA"15 type="cpnatexp: geoAreaCountry">
    <name>Uni ted States</name>
  </broaden>
</assert>
```

Using <assert> in this way can be an advantage if a concept is used more than once in a 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.

### 17.9.9 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 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).

```
<al t I d type="i d type: systemRef" envi ronment="acmesys: mdn acmesys: i i m">SAD02</al t I d>
```

If the DataSet represents a Job ID, the recommended mapping is to the G2 <memberOf> property, a child of <itemMeta> (PCL only):

```
<memberOf type="myref: j obref" l i t e r a l =" SAD02" />
```

### 17.9.10 Headline

Maps to the <headline> child of <contentMeta>, a block type element:

```
<headl i n e xml : l a n g ="en-US">A f i r e f i g h t e r w a l k s p a s t t h e r e m a i n s o f a m i l i t a r y
  j e t t h a t c r a s h e d i n t o h o m e s i n t h e U n i v e r s i t y C i t y n e i g h b o r h o o d o f S a n D i e g o
</headl i n e>
```

Note the use of the @xml:lang property to declare the language and variant “en-US”.

<sup>15</sup> 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.





### 17.9.11 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>
```

### 17.9.12 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>
```

### 17.9.13 Writer/Editor

The caption writer is often a different person to the photographer, so aligns with the 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 24 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="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/
    NewsML-G2_2.4-spec-NewsItem-Power.xsd"
  standard="NewsML-G2"
  standardversion="2.4"
  conformance="power">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_11.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="nat:picture" />
    <provider literal="acmenews.com" />
    <versionCreated>2008-12-09T02:20:00Z</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>
    <facet qcode="fcode:jobtitle">
      <name>Staff Photographer</name>
    </facet>
  </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:DLS" />
  <subject type="cpnat:poi" literal="int001" />
  <headline xml:lang="en-US">A firefighter walks past the remains of a military
jet that crashed into homes in the University City neighborhood of San Diego
</headline>
  <description xml:lang="en-US" role="drol:caption">A firefighter in a flameproof
suit walks past the remains of a military jet that crashed into homes in the
University City neighborhood of San Diego, California December 8, 2008. The
military F-18 jet crashed on Monday into the California neighborhood near San
Diego after the pilot ejected, igniting at least one home, officials said.
</description>
  <creditline>Acme/John Smith</creditline>
  <keyword role="krol:index">us</keyword>
  <keyword role="krol:index">military</keyword>
  <keyword role="krol:index">aviation</keyword>
  <keyword role="krol:index">crash</keyword>
  <keyword role="krol:index">fire</keyword>
  <slugline>USA-CRASH/MILITARY</slugline>
</contentMeta>
<assert literal="int001">
  <type qcode="cpnatexp:geoAreaSublocation" />
  <name>University City</name>
  <broadener qcode="mycity:int002" type="cpnatexp:geoAreaCity">
    <name>San Diego</name>
  </broadener>
  <broadener qcode="mystate:int003" type="cpnatexp:geoAreaProvState">
    <name>California</name>
  </broadener>
  <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:hiRes" size="3764418" contentType="image/jpeg" width="2500"
    height="1445" colourspace="colsp:USWOPv2" />
  </remoteContent>
</contentSet>
</newsItem>

```

## 17.10 Reconciling G2 with embedded metadata

Guidance was discussed in [Reconciling photo metadata](#) for situations where equivalent properties exist in both G2 metadata and embedded (XMP or EXIF) metadata.

The IPTC recommends that embedded administrative metadata, such as <located>, MAY take precedence over G2 metadata, subject to guidance from the provider, and always with caution.

For example, a picture may have GPS metadata embedded by the camera, which may be different from the <located> property entered by a journalist.

The difference may be one of precision: the GPS co-ordinates may be precise, but hardly useful to the ultimate consumer. Even if a human-readable value is derived from the GPS data, would this be better than the information, if accurate, provided by the journalist?

If the difference is due to inaccuracy, the receiver would have no way of knowing whether the journalist has made a mistake, or whether the camera is incorrectly set.



This page intentionally blank

## 18 Advanced Metadata Techniques

### 18.1 Introduction

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.

### 18.2 The Assert wrapper

The Assert wrapper is an optional and repeatable child of any root G2 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 G2 Item and can be referenced by the Item's other G2 properties using either a @qcode or @literal identifier.

The other attributes of <assert> are (from the **editAttributes** group):

Name	Datatype
id	XML Schema ID
creator	QCode
modified	DateOptTime

and from the Internationalization group (**i18nAttributes**):

Name	Datatype
xml:lang	XML Schema language
dir	XML Schema string Enumeration <i>ltr</i> , <i>rtl</i>

The content model for <assert> allows it to contain any element from the IPTC's "NAR" namespace, (and from any another namespace). To avoid ambiguity, the rules that MUST be followed when adding 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.

#### 18.2.1 Business cases

The Assert wrapper has three main uses:

- ❖ To enable a rich inline representation of concepts used in a 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 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 G2 Item instance and are not extracted from a concept in a Concept or Knowledge Item

##### 18.2.1.1 Rich inline representation of concepts

Sometimes it is impractical, inefficient or undesirable to expect receivers of G2 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 G2 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 G2 Item, can be expanded locally using an Assert.

The use of @qcode indicates that the concept is part of a controlled vocabulary, and the concept itself is thus globally available to all G2 Items, whereas a @literal concept identifier is local to the G2 Item in which it is used.

But note that the concept information contained in an Assert in a G2 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 G2 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.

#### 18.2.1.2 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>
  <broaden literal="int002" type="cpnatexp: geoAreaCity">
    <name>Amesbury</name>
  </broaden>
  <broaden literal="int003" type="cpnatexp: geoAreaProvState">
    <name>Wiltshire</name>
  </broaden>
  <broaden qcode="iso3166-1a3: GBR"16 type="cpnatexp: geoAreaCountry">
```

<sup>16</sup> 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 IPTC-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.



```
<name>United Kingdom</name>
</broader>
</assert>
```

#### 18.2.1.2.1 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. Prior to NewsML-G2 2.5 and EventsML-G2 1.4, the only available `<address>` structure was wrapped by `<contactInfo>` with the semantic “how to contact the POI”. It is recommended that the address wrapped by `<contactInfo>` should NOT be used to express the location of the POI.

```
<contentMeta>
  <located qcode="artven:int014" />
  <subject type="cpnat:abstract" qcode="subj:01017001">
    <name>music theatre</name>
  </subject>
  <subject type="cpnat:poi" qcode="artven:int014" />
</contentMeta>
<assert qcode="artven:int014">
  <name>The Metropolitan Opera House</name>
  <definition xml:lang="en-US">
    The Metropolitan Opera House is situated in the Lincoln Center for the
    Performing Arts on the Upper West Side of Manhattan, New York City. <br/>
    The Opera House is located at the center of the Lincoln Center Plaza, at
    the western end of the plaza, at Columbus Avenue between 62nd and
    65th Streets. <br/>
  </definition>
  <POIDetails>
    <address>
      <line>Lincoln Center</line>
      <locality type="geotype:city" literal="int015">
        <name>New York</name>
      </locality>
      <area type="geotype:provstate" literal="int016">
        <name>New York</name>
      </area>
      <country literal="US">
        <name role="nrol:display">United States</name>
      </country>
      <postalCode>10023</postalCode>
    </address>
    <contactInfo>
      <web>http://www.themet.org</web>
    </contactInfo>
  </POIDetails>
</assert>
```

Annotations in the original image:

- Arrows pointing to `int014` in `located` and `subject` attributes: "int014"
- Arrow pointing to `int014` in `assert` attribute: "int014"
- Arrow pointing to `address` element: "location"
- Arrow pointing to `contactInfo` element: "contact info"

#### 18.2.1.3 Using concept details in `<assert>` in previous versions of G2

In 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.

In NewsML-G2 2.5 and EventsML-G2 1.4 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.

##### 18.2.1.3.1 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>

```

#### 18.2.1.3.2 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

```

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/"
  />
  ...

```





## 18.2.2 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 G2](#))

When migrating IIM or IPTC Core (XMP) metadata to 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 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>
```

## 18.3 Inline Reference

The G2 <inlineRef> element complements the <assert> property described above. Whereas <assert> carries supplementary information about concepts referred to in 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 G2 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 <span> 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.


### 18.3.1 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.

### 18.3.2 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>
```



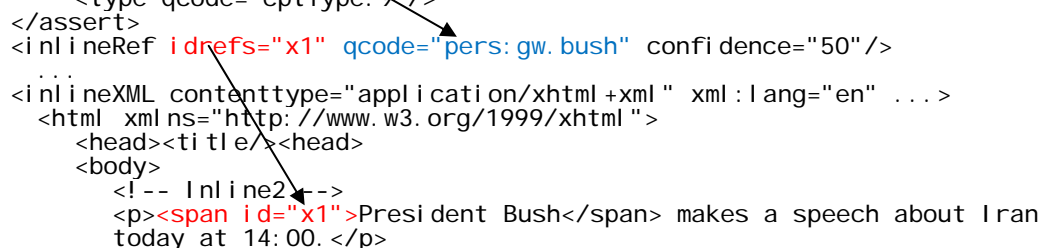
### 18.3.3 Using <inlineRef> and <assert> together

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.

#### 18.3.3.1 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.

### 18.3.3.2 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" facet="geoProp: ci ty"/>
</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`).

### 18.3.3.3 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" rtr: facet="geoProp: 5"/>
</assert>
...
<inlineRef idrefs="x8" qcode="pers: gw. bush" confidence="50"/>
<inlineRef idrefs="x7" qcode="N2: IR" confidence="100"/>
...
```



This page intentionally blank



## 19 Generic Processes and Conventions

### 19.1 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.

### 19.2 Processing Rules for G2 Items

Although at CCL, G2 is designed for maximum inter-operability, providers are strongly recommended to document their implementation of G2 and the provider-specific rules and conventions used.

This information must be provided to receivers so that generic G2 processors can be parameterised accordingly

An obvious example is G2 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.

### 19.3 Publishing Status

The G2 <pubStatus> property uses a mandatory IPTC controlled vocabulary 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 the G2 specification 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 a G2 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.

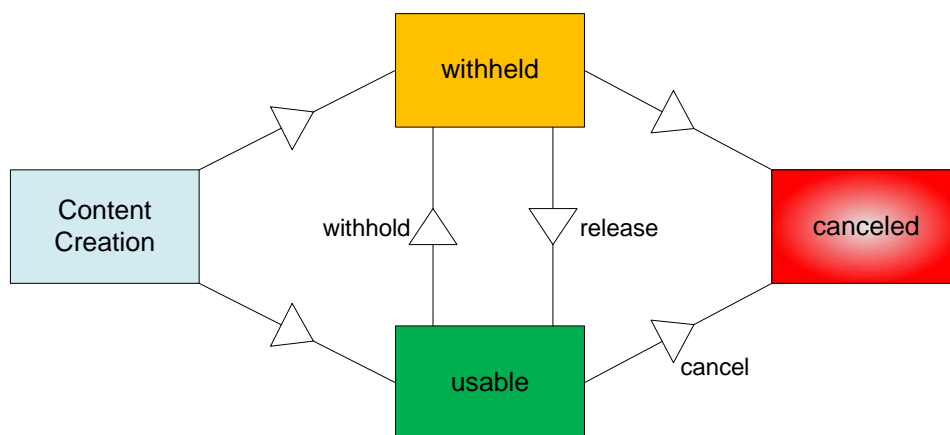


Figure 30: State Transition Diagram for <pubStatus>

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_6.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>
  
```

## 19.4 Embargo

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

Embargo is NOT the same as the Publishing Status. Some systems process the embargo time using software in order to trigger the release of content when the embargo time is passed, but the intention of embargo is also as an information management feature for journalists.

Embargos are generally an unwritten agreement and have no legal force. Their success depends on cooperation between parties not to abuse the system. Possible abuses include imposing unnecessary embargos in order to manage the impact of news<sup>17</sup>, or by breaking embargos and releasing news into the public domain too early.

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.

Up to and including the NewsML-G2 2.2 and EventsML-G2 1.1, `<embargoed>` is defined as the Date and Time (with optional time zone) when an embargo ends, and CANNOT be empty. An embargo time of 12noon GMT on February 9, 2009 would be coded in XML as:

```
<embargoed>2009-02-09T12: 00: 00Z</embargoed>
```

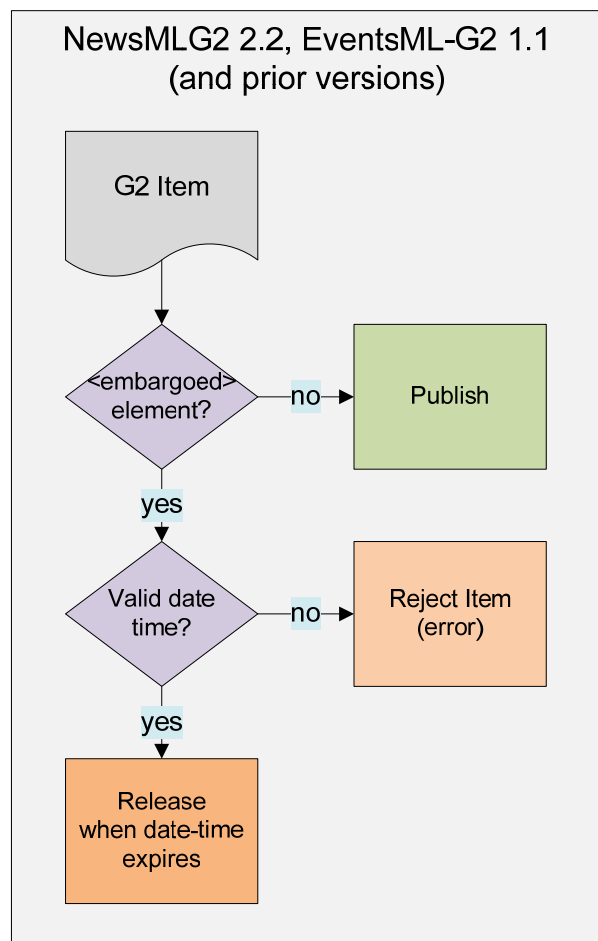


Figure 31: Processing model for `<embargoed>`

<sup>17</sup> 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.



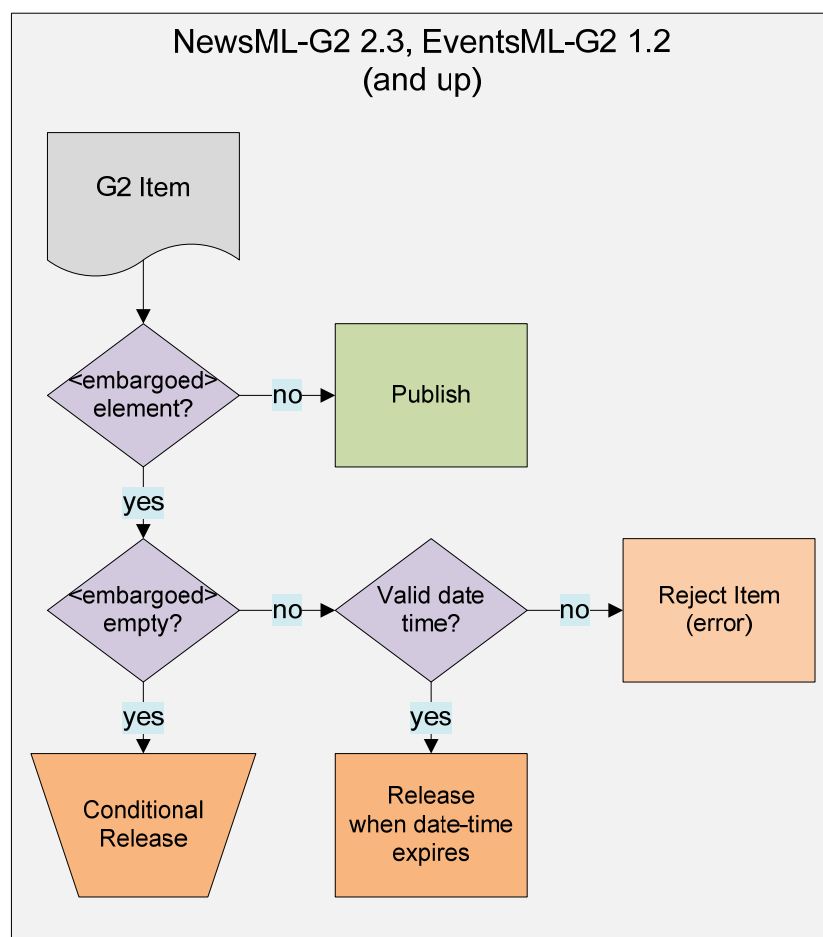
From NewsML-G2 2.3 and EventsML-G2 1.2 the <embargoed> property may be present AND empty. This allows providers to release an item under embargo when the precise date and time that the embargo expires is not known. In these circumstances, an <edNote> or some contractual agreement between the provider 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 G2 Standards Specification Document, available at the IPTC Web site [www.iptc.org](http://www.iptc.org) for further information regarding processing rules for <embargoed>. The difference between the two processing models is illustrated in **Figure 32** below.



**Figure 32: Processing <embargoed> from NewsML-G2 2.3/EventsML-G2 1.2**

## 19.5 Geographical Location

There are two properties of <contentMeta> that express the geographic location(s) associated with a 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 G2 <dateline> property is also available as a human-readable string, for example "MOSCOW, Monday (Reuters)")

The accepted convention, which in some news organisations is formalised as part of a code of practice, is that the Dateline identifies the place where content is created, NOT the place where an event takes place. They may be the same, but this is not necessarily the case.

- ❖ In conflict zones, journalists may not have access to the area where reported events are taking place.
- ❖ Even when physical access is not an issue, journalists may have relied on interviewing people by telephone or on reports from freelance correspondents in order to get the material used to write an article.

The Located property is therefore provided to express the place of origin of content as part of

### Administrative Metadata:

```
<located type="locotype: ci ty" 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>  
  <broadener type="cpnat: geoArea" qcode="stat: 3959">  
    <name>California</name>  
  </broadener>  
  <broadener type="cpnat: geoArea" qcode="country: US">  
    <name>United States</name>  
  </broadener>  
</subject>
```

This news story fragment from Reuters and the accompanying code listing illustrate the use of <located> and <subject>. The geographical subjects of the report are Georgia and South Ossetia, but the report was written in Moscow:

MOSCOW, Monday (Reuters) - The breakaway Georgian region of South Ossetia alleged today that two unexploded Georgian shells landed in its capital Tskhinvali, but Tbilisi dismissed the claim as nonsense.

Both sides have regularly....

### LISTING 25 Illustrating Located, Subject and Dateline

```
< <contentMeta>  
  <contentCreated>2007-02-09T09: 17: 00+03: 00  
  </contentCreated>  
  <located qcode="ci ty: Moscow">  
    <broadener qcode="cntry: Russia" />  
  </located>  
  <creator qcode="web: thomsonreuters. com">  
    <name>Thomson Reuters</name>  
  </creator>
```



```
<language tag="en-US" />
<subject type="cpnat:geoArea" qcode="ci ty: Tskhi nval i ">
  <broader qcode="reg: SouthOsseti a" >
    <name>South Osseti a</name>
  </broader>
</subject>
<subject type="cpnat:geoArea" qcode="ci ty: Tbl i si ">
  <broader qcode="cntry: Georgi a" >
    <name>Georgi a</name>
  </broader>
</subject>
<dateline>MOSCOW, Monday (Reuters)</dateline>
</contentMeta>
```



## 19.6 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 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.

In some content management systems, journalists are able to signal an update to a previous version of an article by issuing a new article, using the same ID, with a flag indicating whether the new article should (for example) replace the previous version, or be appended to it.

G2 allows this facility to be replicated using the <signal> property. Signal uses a QCode to identify an action from a CV. These actions would need to be agreed as part of the provider’s specification for their G2 implementation in order that receivers can correctly process the instruction. The following uses an instruction code “append” from a fictitious “action” scheme to instruct the receiver to append the current item to the previous item (identified by the item GUID)

```
<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="action:append" />
    ....
  </itemMeta>
  ....
</newsItem>
```

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 qcode="action:replace" />
    <link
      rel="irel:previousVersion"
      residref="tag:afp.com,2008:TX-PAR:20090529:JYC80" ←———— previous item
      version="1" /> ←———— first version
    ....
  </itemMeta>
  ....
</newsItem>
```



This page intentionally blank



## 20 Changes to G2-Standards

### 20.1 Introduction

The following is a summary of changes to the G2 Specification since Revision 1 of the G2 Guidelines. Please check the Specification documents available from the IPTC web site ([www.iptc.org](http://www.iptc.org)).

### 20.2 News Architecture (NAR) 1.3 → 1.4

The following changes are inherited by NewsML-G2 2.3 and EventsML-G2 1.2

#### 20.2.1 Revised Embargo

An embargo can now have an undefined date and time. See [Embargo](#)

#### 20.2.2 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](#)

### 20.3 NewsML-G2 v2.2 → v2.3

Inherits the changes described in [20.2](#), plus the following changes specific to NewsML-G2 2.3

#### 20.3.1 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")

#### 20.3.2 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 [7.2.1.10](#)

#### 20.3.3 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 [7.2.2.1](#)

### 20.4 EventsML-G2 v1.1 → v1.2

Inherits the changes described in [20.2](#). No other changes

### 20.5 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>.

## 20.6 News Architecture (NAR) v1.4 → v1.5

The following changes are inherited by NewsML-G2 2.4 and EventsML-G2 1.3

### 20.6.1 @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”, “screen” and “hires”.

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 “hires” renditions in a content set.

### 20.6.2 <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.

### 20.6.3 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.

From NAR 1.5 onwards 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;<sup>18</sup>
- ❖ All other elements MUST be wrapped by their parent element(s), excluding the root element.

### 20.6.4 Scheme Code Encoding

A full processing model for Scheme URIs and QCodes was defined. See [11.9](#)

### 20.6.5 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>

### 20.6.6 Ranking Multiple Elements

Up to NAR 1.5, the elements that support a @rank attribute are:

- ❖ <link>

---

<sup>18</sup> From NewsML-G2 2.5 and EventsML-G2 1.4, the <concept> wrapper is included in this rule, i.e. immediate child properties of <concept> be used directly under the Hint and Extension Point. See example in [16.2.1.2](#)





- ❖ <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.

### 20.6.7 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>
```

### 20.6.8 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.

### 20.6.9 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.

## 20.7 NewsML-G2 v2.3 → v2.4

Inherits the changes described in [20.6](#) plus the following changes specific to NewsML-G2 2.4

### 20.7.1 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
  resi dref="tag: reuters. com, 0000: bi nary_BTRE4A31LE800-THUMBNAIL"
  rendi ti on="rend: thumbnai l"
  contenttype="i mage/j peg"
  format="fmt: j pegBasel i ne"
  wi dth="100"
  hei ght="100"
/>
```

The following example uses explicit dimension units:

```
<remoteContent
  resi dref="tag: reuters. com, 0000: bi nary_BTRE37913MM00-THUMBNAIL"
  rendi ti on="rend: thumbnai l"
  contenttype="i mage/gi f"
  format="fmt: gi f87a"
  wi dth="100" wi dthuni t="di mensi onuni t: poi nts"
  hei ght="100" hei ghtuni t="di mensi onuni t: poi nts"
/>
```

## 20.8 EventsML-G2 v1.2 → v1.3

Inherits the changes described in [20.6](#). No other changes.

## 21 Additional Resources

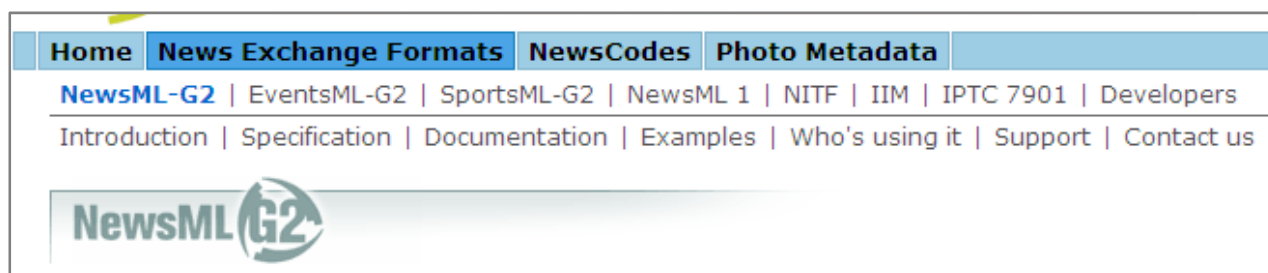
The IPTC web site [www.iptc.org](http://www.iptc.org) has a wealth of resources for implementers. The site is divided into logical areas of interest using horizontal menus under the site banner:



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. The URLs of these packages are also listed below.

Name	Source
G2 Specification Document	The Specification Document is part of a package of documents, examples and XML Schema files obtainable at <a href="http://iptc.org/std/NewsML-G2/NewsML-G2_2.4.zip">http://iptc.org/std/NewsML-G2/NewsML-G2_2.4.zip</a>
NewsML-G2 Examples	See above
EventsML-G2	A package of documents, examples and XML Schema files: <a href="http://iptc.org/std/EventsML-G2/EventsML-G2_1.3.zip">http://iptc.org/std/EventsML-G2/EventsML-G2_1.3.zip</a>
SportsML-G2	A package of documents, examples and XML Schema files may be obtained at <a href="http://iptc.org/std/SportsML/SportsML_2.1.zip">http://iptc.org/std/SportsML/SportsML_2.1.zip</a>
IPTC 7901	A guide to IPTC 7901 is at: <a href="http://iptc.org/std/IPTC7901/1.0/specification/7901V5.pdf">http://iptc.org/std/IPTC7901/1.0/specification/7901V5.pdf</a>
NITF	Resources, including documentation, schema file (version 3.4 only) and DTDs for all versions of NITF are available at <a href="http://iptc.org/std/NITF">http://iptc.org/std/NITF</a>
IPTC4XMP	Resources for IPTC Photo Metadata for XMP may be obtained at the IPTC web <a href="http://iptc.org/photometadata">http://iptc.org/photometadata</a>



This page intentionally blank



## License Agreement

This document is issued under the

### **Non-Exclusive License Agreement for International Press Telecommunications Council Specifications and Related Documentation**

IMPORTANT: International Press Telecommunications Council (IPTC) standard specifications for news (the Specifications) and supporting software, documentation, technical reports, web sites and other material related to the Specifications (the Materials) including the document accompanying this license (the Document), whether in a paper or electronic format, are made available to you subject to the terms stated below. By obtaining, using and/or copying the Specifications or Materials, you (the licensee) agree that you have read, understood, and will comply with the following terms and conditions.

1. The Specifications and Materials are licensed for use only on the condition that you agree to be bound by the terms of this license. Subject to this and other licensing requirements contained herein, you may, on a non-exclusive basis, use the Specifications and Materials.
2. The IPTC openly provides the Specifications and Materials for voluntary use by individuals, partnerships, companies, corporations, organizations and any other entity for use at the entity's own risk. This disclaimer, license and release is intended to apply to the IPTC, its officers, directors, agents, representatives, members, contributors, affiliates, contractors, or co-venturers acting jointly or severally.
3. The Document and translations thereof may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the copyright and license notices and references to the IPTC appearing in the Document and the terms of this Specifications License Agreement are included on all such copies and derivative works. Further, upon the receipt of written permission from the IPTC, the Document may be modified for the purpose of developing applications that use IPTC Specifications or as required to translate the Document into languages other than English.
4. Any use, duplication, distribution, or exploitation of the Document and Specifications and Materials in any manner is at your own risk.
5. NO WARRANTY, EXPRESSED OR IMPLIED, IS MADE REGARDING THE ACCURACY, ADEQUACY, COMPLETENESS, LEGALITY, RELIABILITY OR USEFULNESS OF ANY INFORMATION CONTAINED IN THE DOCUMENT OR IN ANY SPECIFICATION OR OTHER PRODUCT OR SERVICE PRODUCED OR SPONSORED BY THE IPTC. THE DOCUMENT AND THE INFORMATION CONTAINED HEREIN AND INCLUDED IN ANY SPECIFICATION OR OTHER PRODUCT OR SERVICE OF THE IPTC IS PROVIDED ON AN "AS IS" BASIS. THE IPTC DISCLAIMS ALL WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, ANY ACTUAL OR ASSERTED WARRANTY OF NON-INFRINGEMENT OF PROPRIETARY RIGHTS, MERCHANTABILITY, OR FITNESS FOR A PARTICULAR PURPOSE. NEITHER THE IPTC NOR ITS CONTRIBUTORS SHALL BE HELD LIABLE FOR ANY IMPROPER OR INCORRECT USE OF INFORMATION. NEITHER THE IPTC NOR ITS CONTRIBUTORS ASSUME ANY RESPONSIBILITY FOR ANYONE'S USE OF INFORMATION PROVIDED BY THE IPTC. IN NO EVENT SHALL THE IPTC OR ITS CONTRIBUTORS BE LIABLE TO ANYONE FOR DAMAGES OF ANY KIND, INCLUDING BUT NOT LIMITED TO, COMPENSATORY DAMAGES, LOST PROFITS, LOST DATA OR ANY FORM OF SPECIAL, INCIDENTAL, INDIRECT, CONSEQUENTIAL OR PUNITIVE DAMAGES OF ANY KIND WHETHER BASED ON BREACH OF CONTRACT OR WARRANTY, TORT, PRODUCT LIABILITY OR OTHERWISE.
6. The IPTC takes no position regarding the validity or scope of any Intellectual Property or other rights that might be claimed to pertain to the implementation or use of the technology described in the Document or the extent to which any license under such rights might or might not be available. The IPTC does not represent that it has made any effort to identify any such rights. Copies of claims of rights made available for publication, assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of the Specifications and Materials, can be obtained from the Managing Director of the IPTC.
7. By using the Specifications and Materials including the Document in any manner or for any purpose, you release the IPTC from all liabilities, claims, causes of action, allegations, losses, injuries, damages, or detriments of any nature arising from or relating to the use of the Specifications, Materials or any portion



thereof. You further agree not to file a lawsuit, make a claim, or take any other formal or informal legal action against the IPTC, resulting from your acquisition, use, duplication, distribution, or exploitation of the Specifications, Materials or any portion thereof. Finally, you hereby agree that the IPTC is not liable for any direct, indirect, special or consequential damages arising from or relating to your acquisition, use, duplication, distribution, or exploitation of the Specifications, Materials or any portion thereof.

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

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

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

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

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

13. This Specifications License Agreement is perpetual subject to your conformance to the terms of this Agreement. The IPTC may terminate this Specifications License Agreement immediately upon your breach of this Agreement and, upon such termination you will cease all use, duplication, distribution, and/or exploitation in any manner of the Specifications and Materials.

14. This Specifications License Agreement reflects the entire agreement of the parties regarding the subject matter hereof and supersedes all prior agreements or representations regarding such matters, whether written or oral. To the extent any portion or provision of this Specifications License Agreement is found to be illegal or unenforceable, then the remaining provisions of this Specifications License Agreement will remain in full force and effect and the illegal or unenforceable provision will be construed to give it such effect as it may properly have that is consistent with the intentions of the parties.

15. This Specifications License Agreement may only be modified in writing signed by an authorized representative of the IPTC.

16. This Specifications License Agreement is governed by the law of United Kingdom, as such law is applied to contracts made and fully performed in the United Kingdom. Any disputes arising from or relating to this Specifications License Agreement will be resolved in the courts of the United Kingdom. You consent to the jurisdiction of such courts over you and covenant not to assert before such courts any objection to proceeding in such forums.

**IF YOU DO NOT AGREE TO THESE TERMS YOU MUST CEASE ALL USE OF THE SPECIFICATIONS AND MATERIALS NOW.**

**IF YOU HAVE ANY QUESTIONS ABOUT THESE TERMS, PLEASE CONTACT THE MANAGING DIRECTOR OF THE INTERNATIONAL PRESS TELECOMMUNICATION COUNCIL.**

**AS OF THE DATE OF THIS REVISION OF THIS SPECIFICATIONS LICENSE AGREEMENT YOU MAY CONTACT THE IPTC at <http://www.iptc.org> .**

License agreement version of: 30 January 2006