



IPTC Standards



Specification Version 2.23
Power Conformance Level

Document Revision 1



Copyrights and License

Copyright © 2017 by IPTC, the International Press Telecommunications Council. All Rights Reserved.

The IPTC NewsML-G2/EventsML-G2 specification document is published under the Creative Commons Attribution 4.0 license - see the full license agreement at <http://creativecommons.org/licenses/by/4.0/>. By obtaining, using and/or copying this document, you (the licensee) agree that you have read, understood, and will comply with the terms and conditions of the license.

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

Acknowledgments

This documentation is the result of a team effort by members of the International Press Telecommunications Council, with input and assistance from other contributors.

The effort to develop NewsML-G2 was led by Laurent Le Meur (Agence France-Presse) and these persons contributed (ordered by family name): Yannick Beynet (Agence France-Presse), Mark Birbeck (x-port.net Ltd.), Dave Compton (Reuters), Jay Cousins (RivCom), Jean-Pierre Evain (EBU), John Evans (Transtel), Takahiro Fujiwara (EAST Co. Ltd.), Andreas Gebhard (Getty Images), Philipp Gortan (APA), Darko Gulija (HINA), Paul Harman (Press Association), Gerald Innerwinkler (APA), Johan Lindgren (Tidningarnas Telegrambyrå), Jayson Lorenzen (BusinessWire), Philippe Mougin (AFP), Stuart Myles (initially Dow Jones, now Associated Press), Kalle Rathje (dpa), Robert Schmidt-Nia (dpa), Michael Steidl (IPTC Managing Director), Ulf Wingstedt (CNET), Misha Wolf (Reuters).

The Technical Writer of the initial version of the document was Scott Meltzer, now it is maintained by Michael Steidl.

About the Standards

Specification Versioning History

Version	Date	Approved by	Remarks
2.0	2008-01-31	IPTC Standards Committee	NewsML-G2 approval
2.7	2010-06-30	IPTC Standards Committee	NewsML-G2 approval
1.6	2010-06-30	IPTC Standards Committee	EventsML-G2 approval
2.9	2011-06-09	IPTC Standards Committee	joint NewsML-G2/EventsML-G2 approval
2.12	2012-06-13	IPTC Standards Committee	NewsML-G2 including EventsML-G2
2.15	2013-06-26	IPTC Standards Committee	NewsML-G2 including EventsML-G2
2.18	2014-06-18	IPTC Standards Committee	NewsML-G2 including EventsML-G2
2.21	2015-06-03	IPTC Standards Committee	NewsML-G2 including EventsML-G2
2.23	2016-06-15	IPTC Standards Committee	NewsML-G2 including EventsML-G2

The specifications of NewsML-G2 and EventsML-G2 have been published separately up to the standard versions EventsML-G2 1.7 and NewsML-G2 2.8. As the design and a vast majority of the specified structures are shared between both standards the IPTC decided in June 2011 to merge the specifications into a single document and to provide all specifications and other documentation under the main branding NewsML-G2 and in the NewsML-G2 folders of the IPTC web server, see below Status of this Document.

This step has no impact on the structure of EventsML-G2 or NewsML-G2.



Document Revision History

Revision	Issue Date	Author (revisedby)	Remark
1	2017-02-21	Michael Steidl	

About this Document

This document specifies the IPTC news exchange standard NewsML-G2 and its event focused sibling EventsML-G2 which is a conceptual and processing model making freely available the IPTC knowledge of the most effective ways to structure, describe, manage and exchange general news and event data.

Status of this Document

This document is under the governance of the IPTC News Exchange Formats Working Party and its sub-groups NewsML-G2 Working Group and EventsML-G2 Working Group.

This is a specification document which was endorsed by the IPTC members and may be updated, replaced or obsoleted by other documents at any time.

Public versions of this document and of related IPTC documents are available in subdirectories of

<http://www.iptc.org/std/NewsML-G2/2.23/>

Public comments should be sent to the forum and mailing list at:

<http://tech.groups.yahoo.com/group/newsml-g2>

A page with all errata not covered by the latest version of the NewsML-G2 specification is available at:

<http://www.iptc.org/goto?NewsML-G2-Errata>

The Full Set of Specification Documents

The full set of specification documents for NewsML-G2 2.23 consists of (file names are added, # is to be substituted by the most current document revision number):

This Specification document - NewsML-G2_2.23-spec-PCL_#.pdf

XML Schema files applicable to the Core Conformance Level (see [Conformance Levels](#)):

- ◆ NewsML-G2_2.23-spec-All-Core_#.xsd

XML Schema files applicable to the Power Conformance Level (see [Conformance Levels](#)):

- ◆ NewsML-G2_2.23-spec-All-

Power_#.xsd All files above can be obtained

from:

<http://www.iptc.org/std/NewsML-G2/2.23/specification/>

Note on the XML Schema File Names

XML Schemas are revised for two reasons:

- ◆ The NewsML-G2 specifications have been changed: this results in a new version of the standard, this will be reflected by a new path to files and a new standard version number like NewsML-G2_2.23.
- ◆ The XML Schema has been edited a) to fix errors and b) to change non-normative parts, like the wording of an element's annotation: this is reflected by a new revision number at the end of the file name like the "3" in NewsML-G2_2.23-spec-Framework-Core_3.xsd.

The XML Schema files without the document revision number (e.g. "_3") at the end of the file name are true copies of the latest document revision. This allows applying a persistent reference to the latest XML Schema file version regardless of any edits of the document.

Terminology

This document uses the terms MUST (NOT), SHOULD (NOT) and MAY as defined in [RFC2119].



Table of Contents

Copyrights and License.....	2
Acknowledgments.....	2
About the Standards	2
14.3 Specification Versioning History	2
14.4 Document Revision History	3
14.5 Status of this Document.....	3
14.6 The Full Set of Specification Documents	3
14.7 Note on the XML Schema File Names	3
14.8 Terminology	3
1 Introduction to NewsML-G2.....	7
1.1 History.....	7
1.2 Conformance Levels.....	7
1.3 Supporting documents.....	7
2 Representing News - newsItem	8
2.1 Description.....	8
2.2 Indication of Compliance with a Standard and Conformance Level.....	8
2.3 Identification and Versioning.....	8
2.4 Catalog of Controlled Vocabularies	9
2.5 Signature Information	10
2.6 Rights Information	10
2.7 Item Metadata.....	16
14.9 2.8 Item Links.....	19
2.9 News Content Metadata	20
2.10 Part of Content Metadata	22
2.11 Assertions About Concepts	23
2.12 References to Inline Concepts.....	23
2.13 Document Derivation of Concepts	24
2.14 newsItem Content.....	24
2.15 News Content Characteristics	24
2.16 Channels	25
3 Introduction to EventsML-G2	26
3.1 Overview.....	26
3.2 Definitions	27
3.3 EventsML-G2 and iCalendar.....	27
4 Events 30	
4.1 The Core Information about Events	30
4.2 Event Information in Items.....	31
5 Representing Concept Information - concept Component	34
5.1 Concept Component.....	34
5.2 Relationships Between Concepts.....	35
5.3 Details Associated with Specific Entities.....	35
6 Managing Individual Concepts - conceptItem.....	37
6.1 Description.....	37
6.2 Structure of a Concept Item	37
6.3 Item Class.....	37
6.4 Concept related Metadata	37
6.5 Metadata Helpers.....	38
6.6 conceptItem Content	38
7 Managing Sets of Concepts - knowledgeItem.....	39
7.1 Description.....	39
7.2 Structure of a Knowledge Item	39
7.3 Item Class.....	39
7.4 Knowledge Related Metadata.....	39
7.5 Metadata Helpers.....	40
7.6 knowledgeItem Content	40
8 Packaging Items - packageItem.....	41
8.1 Structure of a Package Item	41
8.2 Item Class.....	41
8.3 Package Related Metadata	41



8.4	Metadata Helpers.....	41
8.5	packageItem Content.....	41
9	Planning news coverage - planningItem.....	43
9.1	Description.....	43
9.2	Structure of Planning Item.....	43
9.3	Item Class.....	43
9.4	PlanningRelatedMetadata.....	43
9.5	Metadata Helpers.....	43
9.6	Planning ItemContent	44
9.7	Processing Considerations.....	44
10	Dealing with Controlled Values.....	44
10.1	{scheme, code} Pair, Scheme URI and Concept URI.....	44
10.2	Qualified Code (QCode).....	45
10.3	Processing Catalogs.....	47
10.4	Processing Schemes	48
10.5	Qualified and Typed Properties.....	49
10.6	FlexibleProperties	49
10.7	Composite Concepts	50
10.8	EditingAttributes.....	50
11	Managing Catalogs - catalogItem	52
11.1	Description.....	52
11.2	Structure of a Catalog Item.....	52
11.3	Item Class.....	52
11.4	Catalog Related Metadata.....	52
11.5	Catalog Item Content.....	52
12	Dealing with Labels and Blocks.....	53
12.1	Introduction.....	53
12.2	InternationalizationAttributes.....	53
13	Exchanging Items - newsMessage.....	54
13.1	Description.....	54
13.2	Message Information.....	54
13.3	About Using Schemes in a newsMessage	54
14	Specification Reference.....	55
14.1	Introduction to the Common Components.....	55
14.2	General Specifications	55
14.3	Implementation Design Rules.....	56
14.4	Processing Model Terminology	56
14.5	Element specifications.....	58
14.5.1	Accountable Person – accountable.....	58
14.5.2	Action in Hop History – action.....	58
14.5.3	Alternative Identifier – altId	58
14.5.4	Alternative Locator - altLoc	58
14.5.5	Alternative Representation - altRep	58
14.5.6	Assertion - assert.....	58
14.5.7	Bag Item – bit	58
14.5.8	Broader – broader	59
14.5.9	By - by	59
14.5.10	Channel for News Message – channel	59
14.5.11	Circle – circle.....	59
14.5.12	Concept Definition – definition.....	59
14.5.13	Concept Name – name	59
14.5.14	Confirmation – confirmation.....	59
14.5.15	Contact Information – contactInfo.....	59
14.5.16	Provider – provider	59
14.5.17	Date Content Created - contentCreated.....	60
14.5.18	Date Content Modified – contentModified.....	60
14.5.19	Date of Transmission – sent.....	60
14.5.20	Dateline – dateline.....	60
14.5.21	Description – description.....	60
14.5.22	Destination – destination	60
14.5.23	Editorial Service – service	60
14.5.24	Event – event	60
14.5.25	Events – events	60
14.5.26	G2 Content Type – g2Contenttype.....	60
14.5.27	Generator Tool – generator.....	60
14.5.28	Geographic Position – position.....	61
14.5.29	Group – group.....	61



14.5.30	Has Financial Instrument – hasInstrument.....	61
14.5.31	Hierarchy Info – hierarchyInfo.....	61
14.5.32	Hop – hop.....	61
14.5.33	Inline Data – inlineData.....	61
14.5.34	Instant Messaging Address – im.....	61
14.5.35	Information Source – infoSource.....	62
14.5.36	Item Class – itemClass.....	62
14.5.37	Item Set - itemSet.....	62
14.5.38	Keyword – keyword.....	62
14.5.39	Language – language.....	62
14.5.40	Line (of geoArea) – line.....	62
14.5.41	Link – link.....	62
14.5.42	Located – located.....	63
14.5.43	News Message Header – header.....	63
14.5.44	News Content Characteristics – newsContentCharacteristics.....	63
14.5.45	News Coverage (of Concept Item) – newsCoverage.....	63
14.5.46	NewsCoverage (of Planning Item) – newsCoverage.....	63
14.5.47	News Coverage Status – newsCoverageStatus.....	63
14.5.48	Origin – origin.....	63
14.5.49	Organiser – organiser.....	64
14.5.50	Participant – participant.....	64
14.5.51	Phone Number – phone.....	64
14.5.52	Polygon – polygon.....	64
14.5.53	Postal Address.....	64
14.5.54	Postal Address of a Point of Interest – address (in POI structure).....	64
14.5.55	Profile – profile.....	64
14.5.56	Publish Status – pubStatus.....	64
14.5.57	Rating – rating.....	64
14.5.58	Recurrence Group.....	64
14.5.59	Registration - registration.....	64
14.5.60	Remote Content – remoteContent.....	64
14.5.61	Role in the Content Stream – role.....	65
14.5.62	Role in the Workflow – role.....	65
14.5.63	Ruby – ruby.....	65
14.5.64	SameAs for a Scheme - sameAs.....	65
14.5.65	SameAs Scheme – sameAsScheme.....	65
14.5.66	Sender – sender.....	65
14.5.67	Signal – signal.....	65
14.5.68	Slugline – slugline.....	65
14.5.69	Time Delimiter – timeDelim.....	65
14.5.70	Transmission Identifier – transmitId.....	66
14.5.71	Usage Terms – usageTerms.....	66
14.6	Datatype specifications.....	66
14.6.1	ApproximateDateTimePropType.....	66
14.6.2	AudienceType.....	66
14.6.3	BlockType.....	66
14.6.4	ConceptIdType.....	66
14.6.5	DateOptTimePropType and DateOptTimeType.....	66
14.6.6	Flex...PropType (multiple).....	67
14.6.7	Label1Type.....	67
14.6.8	Link1Type.....	67
14.6.9	RecurrenceRuleType.....	67
14.6.10	TruncatedDateTimePropType and TruncatedDateTimeType.....	67
14.7	Attribute (Group) Specifications.....	67
14.7.1	Internationalization Attributes - i18nAttributes.....	67
14.7.2	Ranking Attributes - rankingAttributes.....	67
14.7.3	Quantify Attributes - <i>quantifyAttributes</i>	68
14.7.4	Orientation Attribute - origination.....	68
15	Glossary.....	70
16	References.....	73
	Contact Information.....	73



1 Introduction to NewsML-G2

NewsML™ is a media-independent news exchange format for general news.

News exchange is a method of moving around not only the core news content, but also data that describe the content in an abstract way (i.e. metadata), information about how to handle news in an appropriate way (i.e. news management data), information about the packaging of news information, and finally information about the technical transfer itself.

1.1 History

The initial version of NewsML, version 1.0, was approved in October 2000. Since then it went along with minor revisions: version 1.1 was approved in October 2002; version 1.2 was approved in October 2003.

In 2004, the user-experience with NewsML was evaluated by IPTC, and it was decided to create a consistent set of complementary standards as a comprehensive and inter-operable way to move all types of data between media systems in order to make news exchange efficient and reliable. This set of standards is now the IPTC family of G2-Standards, it includes NewsML-G2, EventsML-G2 and SportsML-G2; NewsML-G2 is the brand name for all of them.

The family of IPTC G2-Standards is built on a common structural and function framework called the IPTC News Architecture (NAR). For this reason many components are common across the members of the G2-Standards.

To better understand the terminology used in the G2-Standards specifications we recommend the [Glossary](#) as a reference, as it provides an extensive set of terms and their definitions.

Since the initial release of NewsML-G2 in 2008 many news providers have adopted this standard and IPTC has extended and slightly modified the specifications by raised change requests.

To reflect implementations IPTC has conducted a survey of the properties used in practice in 2013 and the resulting “mainstream profile” is shown in chapter 19 of the G2 Implementation Guide, see below in 1.3.

1.2 Conformance Levels

Different conformance levels are defined in the model, each of them related to a level of complexity (at the conceptual and processing level) of the related Items. This feature adds modularity to the model.

The current model defines two conformance levels named Core Conformance Level (CCL) and Power Conformance Level (PCL). The Core Conformance Level is focused on simplicity and interoperability. The Power Conformance Level is a superset of the Core Conformance Level which gives more flexibility to providers who choose it, but the recipient processors are more complex to program to comply with PCL and interoperability is lower than for CCL as not all recipients will implement the power level.

A compliant processor must therefore assert supporting either Core or Power functionality.

As the Power features are only an extension of the Core features, a Core compliant processor **SHOULD** process Power Items by simply ignoring the information pertaining to the Power Conformance Level.

1.3 Supporting documents

This document and the XML Schema files are the formal specification of the G2 Standards.

IPTC also provides documents supporting the implementation of the standards in subfolders of <http://www.iptc.org/std/NewsML-G2/2.23/>

- ◆ G2 Quick Start Guides: small documents explaining how to take the first steps for successfully start with text, photos, video and packages thereof using NewsML-G2
- ◆ The G2 Implementation Guide: a comprehensive guideline covering also special work areas like the management of controlled vocabularies and migrating from existing standards to NewsML-G2.
- ◆ A set of almost 30 NewsML-G2 and EventsML-G2 example XML document.
- ◆ A Structure Matrix table showing for each property its attributes.



2 Representing News - `newsItem`

An XML Schema file corresponding to the specifications for this item is available (see [The Full Set of Specification Documents](#) on page 3).

2.1 Description

A `newsItem` aims to convey news with the sense of the reporting of a newsworthy event or fact. Its content is gathered by journalists, presented with a journalistic style, and updated according to the progression of the story.

Examples of `newsItems` are a news report, a picture, a graphical illustration of some event, a video clip or an illustrated biography.

Typical characteristics of a `newsItem` are:

- ◆ Its content may be of any media type or format, e.g., the thumbnail, preview and high definition renditions of a picture.
- ◆ It can also convey more structured news information, e.g., information about companies, sports events and general events, in instances when this information is related to an event or fact.
- ◆ Its content is of short term interest: `newsItems` are volatile, and interest in them fades as time passes ("nothing is older than yesterday's news").
- ◆ It is expressed via a set of alternative renditions of some media content.
- ◆ It will usually be updated only for a short period of time, as long as the covered event evolves, and then may be archived.
- ◆ It refers to an arbitrary set of concepts and entities.
- ◆ It may be associated with other `newsItems` or Web resources via typed links.

2.2 Indication of Compliance with a Standard and Conformance Level

The IPTC `newsItem` *standard* attribute MUST be set to "NewsML-G2" from NewsML-G2 2.9 on. "EventsML-G2" MAY be used up to version 1.7 of this standard.

The schema version to which the `newsItem` conforms MUST be indicated as an attribute. The current version is identified by the string "2.23".

The IPTC conformance level to which the `newsItem` conforms in this specification MUST be indicated by the conformance attribute value "power".

Sample Core Conformance Level:

```
<newsItem standard="NewsML-G2"
  standardversion="2.23"
  xmlns="http://iptc.org/std/nar/2006-10-01/" >
</newsItem>
```

Sample Power Conformance Level:

```
<newsItem standard="NewsML-G2" standardversion="2.23"
  conformance="power" xmlns="http://iptc.org/std/nar/2006-10-01/" >
</newsItem>
```

2.3 Identification and Versioning

It is possible to positively identify a `newsItem` as it moves through the news workflow and is transferred from place to place and from system to system.

A `newsItem` MUST have a *guid* attribute holding a persistent and globally unique identifier. IPTC recommends using an IRI but this is not a requirement. Any string capable of acting as a globally unique identifier may be used.



Note: The IPTC will provide the newsml-URN for this purpose, specified by a successor of RFC-3085.

A newsItem MAY have a **version** attribute, and this version MUST be incremented when the content of the Item is updated. The first version MUST be numbered 1: if the version is not explicitly set, this value must be assumed by the recipient of the Item.

The **standardversion** attribute must reflect the version of the standard as it is implemented by the corresponding XML Schema.

Sample:

```
<newsItem standard="NewsML-G2" standardversion="2.23"
  conformance="power" guid="urn:newsml:iptc.org:20071231:sample"
  version="2" xmlns="http://iptc.org/std/nar/2006-10-01/" >
</newsItem>
```

2.4 Catalog of Controlled Vocabularies

NewsML-G2 recommends the use of controlled values for most properties. Each news provider is free to use their own taxonomies of subjects, genres, geopolitical areas, organisations etc., and to use any value scheme it decides in the Items it provides. A provider must therefore indicate the list of the schemes he is using.

Cataloguing information MUST be included at the top of each Item.

A **catalog** is defined as a set of scheme declarations in use by a news provider for a given Item.

Due to the large number of the same schemes potentially used in many single Items, and knowing that bandwidth is very important to the News industry, the catalog may be stored remotely and referenced by the Item.

A remote catalog MUST have a **href** attribute which contains the URL of a remote catalog. A remote catalog takes the form of an XML file with a catalog element as root. (An XML requirement is to add the NewsML-G2 namespace definition to the catalog element.)

The URL of a remote catalog acts both as a locator and a global identifier, therefore:

- ◆ The URL of a remote catalog MUST NOT be relative.
- ◆ If a remote catalog is functionally changed, the URL used to access it MUST be changed. Functional changes are:
 - the addition or removal of a scheme declaration,
 - a change to any of the scheme aliases,
 - a change to any of the scheme URIs.
 - a change to any of the combinations of schema alias and scheme URI.

One or more additional titles (see **title {itemMeta}**) for a catalog or catalogRef MAY be provided in different languages and variants.

To extend the information about the catalog some optional attributes of the catalog element may be used:

- url: defines the location of the catalog as remote resource.
- authority: defines the authority controlling this catalog
- guid: a Globally Unique Identifier for this kind of catalog as managed by a provider
- version: version corresponding to the guid of the catalog

As some required properties of any G2 Item take a QCode as a value, at least one catalog or remoteCatalog MUST be present.

In general, a given provider will define a unique catalog of all used schemes, store it in a central repository and reference it from all Items it provides. A provider MAY declare several catalogs in the same Item. This may be especially useful for an aggregator which uses property values from different sources, but

requires a way to avoid scheme alias clashes. In this case, catalog and remote catalog elements MAY appear in any order, and their order is not relevant.



The main reason for using a **sameAsScheme** indicator for a scheme in the catalog is speeding up QCode processing: a G2 processor does not have to check the individual concept for its sameAs relationships but can apply this relationship directly to a concept if the scheme identifier of this concept (used as property value) matches the scheme identifier in the sameAsScheme child in the catalog.

Another reason for establishing a sameAsSchema relationship between a scheme A of a provider and a referenced scheme B is to provide additional information about concepts, this could be identical information from scheme B in a different language or deeper information in the same language(s) as available with scheme B.

Detailed information on the structure of catalogs and their processing is given in **Dealing with Controlled Values ()**.

Sample:

```
<newsItem standard="NewsML-G2" standardversion="2.15"
  guid="urn:newsml:iptc.org:20071231:sample" version="2" xmlns="http://
  iptc.org/std/nar/2006-10-01/" >
  <catalogRef href="http://aprovider.com/cv/newsml-g2-catalog-4.xml"/>
  ....
</newsItem>
```

2.5 Signature Information

A digital signature may be associated with a whole Item or only parts of it. For example, it is possible to sign each individual news content component of a newsItem using their local identifiers as a local reference.

A digital signature is a unique seal placed on data. It is very difficult to forge and assures that any change made to the signed data cannot go undetected.

This specification supports the model and syntax defined by the W3C in [XMLDSIG], and introduced by the following: "XML Signatures provide integrity, message authentication, and/or signer authentication services for data of any type, whether located within the XML that includes the signature or elsewhere".

This specification model excludes two functionalities defined by the W3C XML-Signature Processing Recommendation. These are: "Signed content included within an XML Signature Construct" and "Detached Signatures".

Therefore this specification offers the following features:

- ◆ A Signature **MUST** be "enveloped" (the Signature Component is contained within the Item being signed).
- ◆ A Signature **MUST** sign the Item containing the Signature component or child components of the Item containing the Signature.
- ◆ The Signature **MUST NOT** be "enveloping" (it cannot sign content found within the signature itself).
- ◆ A Signature **MUST NOT** be "detached" (a detached Signature Component would not be contained within the Item being signed and could be external to the containing document).
- ◆ A Signature **MUST NOT** be related to Items and Components external to the enclosing document (via references).

2.6 Rights Information

The content of a newsItem is bound to a set of copyrights and licensing information.

A **rightsInfo** wrapper element acts as a container for a set of properties related to rights, which offer a basic expression of the copyright and usage conditions associated with an Item.

This set is limited to an **accountable** person, a **copyrightHolder** and a set of **copyrightNotice** elements and **usageTerms**.

The order of the properties is flexible: The non-repeatable properties **MUST** come first, then the repeatable properties **MAY** be inserted in any order.

The expression of rights can be verbose, and the volume of information exchanged or stored may suffer from

the repetition of such information. Therefore each property provides an *href* attribute as an alternative locator of a remote expression of rights. In the case where both inline and remote expression of rights is indicated, the inline expression **MUST** take precedence.

In some situations, different parts of the content are associated with different sets of rights; the rightsInfo element is therefore repeatable.

Each set of rights provides a set of optional attributes (idrefs, scope, aspect), which indicate which part of the content is bound to these rights. Please review the comprehensive Processing Model below.

The rightsInfo element also provides optional time validity attributes (*validfrom* and *validto*) which express the date and time between which the set of rights properties apply.

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

2.6.1 Processing Model

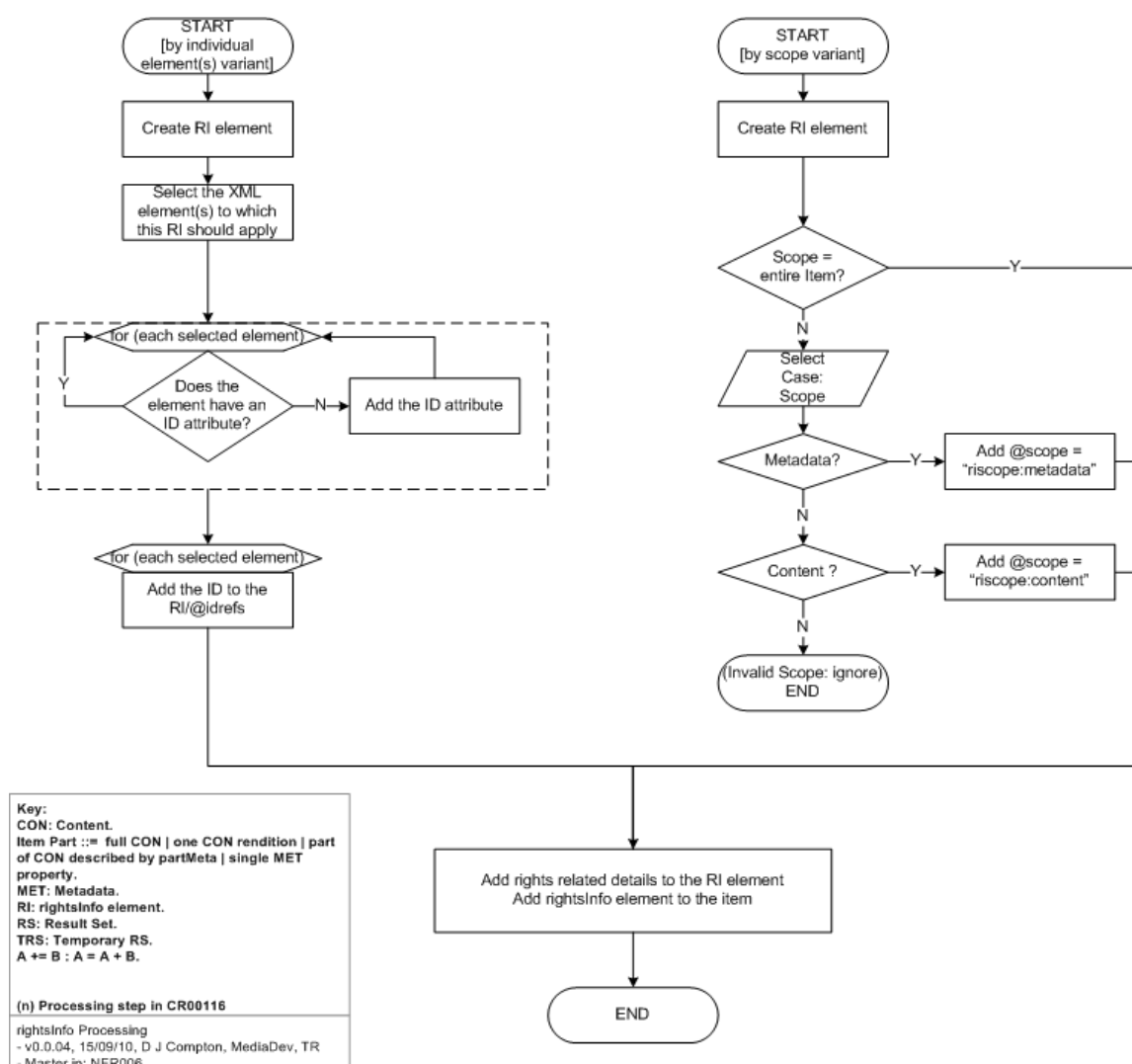
Rules for adding rightsInfo expressions - taking the News Provider View:

To be answered: Which markup does a provider has to apply?

Use Case 1 / 3:

Rules for adding <rightsInfo> expressions – taking the News Provider View:

To be answered: How to apply rightsInfo elements referencing only a fraction of a G2 Item?



1. How a rightsInfo element applies to an Item can be refined in two ways: a) making a statement about



the scope, i.e. whether this rightsInfo element applies to the whole or part(s) of the Item, and b) making a statement about the rights-related aspect of the Item or part(s) of the Item to which rightsInfo applies.

2. There are two ways to express the scope:

2.1. In a general way: all elements of an Item are split into a) the set of metadata properties and b) the content. Thus it can be expressed that

- rightsInfo is about the Item as a whole by not having a @scope attribute
 - rightsInfo is about the metadata properties only by adding a @scope attribute with a value of "riscope:metadata"
 - rightsInfo is about the content only by adding a @scope attribute with a value of "riscope:content"
- To see which parts of an Item fall under the content-scope, and which parts under the metadatascope, check the definition in the Rights Info Scope NewsCodes. When making a statement about the scope in this general way an @idrefs attribute MUST NOT be present on this rightsInfo element (else the scope will only apply to the element(s) with a corresponding @id).

2.2. In a specific way: by adding the ID(s) of XML element(s) to the @idrefs attribute this rightsInfo applies only to all element(s) which have a corresponding @id.

This specific addressing of elements overrides rightsInfo expressions which use the mechanism described in 2.1 above.

The application of rightsInfo is not inherited by the children of itemMeta and contentMeta if these wrapper elements are targeted using their IDs. Therefore their IDs should not be added to @idrefs. If the referenced XML element is a partMeta element then:

- If a @scope attribute is not present then rightsInfo applies to both the content described by this partMeta element and to the metadata children of this partMeta element.
- If a @scope attribute is present its value(s) determines whether rightsInfo applies to the content described by this partMeta element or to the metadata children of this partMeta element.

In compliance with the specification of the @idrefs attribute, IDs of only the following XML elements may be included into the list of values of @idrefs:

- all metadata properties as per the definition of the Rights Info NewsCode for "riscope:metadata".
- the child elements inlineXML, inlineData and remoteContent of contentSet of a News Item as they provide renditions of the full content, the child element concept of conceptSet of a Knowledge Item and the child element group of groupSet of a Package Item.

Explicitly excluded are all child elements of inlineXML of a News Item as they contain only parts of the content. In this case a partMeta element must be used to describe this part and the value of the @partid attribute of this partMeta element must be added to the list of values of the @idref attribute of the rightsInfo element.

3. The @scope and @idrefs attributes allow one to determine to which XML elements a rightsInfo element applies. In some cases it is necessary to associate a rightsInfo element with a particular aspect of an XML element. For example, a keyword element may contain a term associated with a photograph. One aspect of the keyword element to which a rightsInfo element may apply is the term itself. Another aspect to which a rightsInfo element may apply is the selection and application of this term to this photograph. Rights on these two aspects could be different. The @aspect attribute allows one to determine to which rights-related aspects the rightsInfo element applies.

If an @aspect attribute is not present then all aspects from the Rights Aspect NewsCodes apply.

If an @aspect attribute is present then only the aspects from the Rights Aspect NewsCodes listed in the attribute apply.

If a target does not support a specific aspect which is listed in the @aspect attribute then this aspect should be ignored for this target.

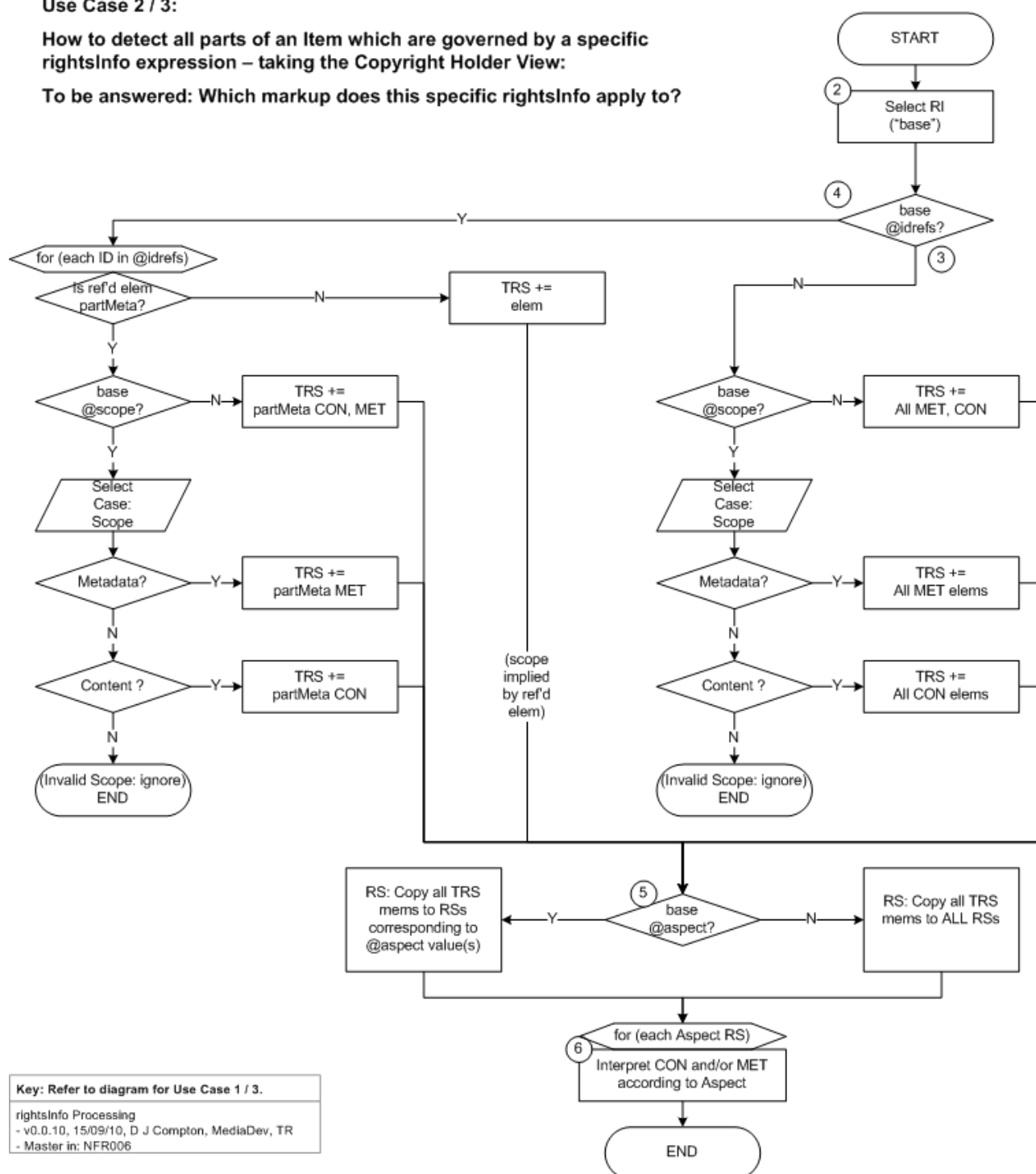
How to detect all parts of an Item which are governed by a specific rightsInfo expression - taking the Copyright Holder View:

To be answered: Which markup does this specific rightsInfo apply to?

Use Case 2 / 3:

How to detect all parts of an Item which are governed by a specific rightsInfo expression – taking the Copyright Holder View:

To be answered: Which markup does this specific rightsInfo apply to?



1. The goal of the processing: the result will be multiple sets of elements and/or parts of content which all are governed by a rightsInfo expression. Each of the sets a) corresponds to one of the RightsAspect NewsCodes, and b) may be empty after the processing if no corresponding parts of an item were found.

2. Select the rightsInfo element to be processed; this is the "base" for all subsequent processing steps.

3. If no @idrefs attribute exists in the base:

3.1. If a @scope attribute is not present: all the content and all metadata properties of this item are governed by the base's rights expression; they all should be included into a temporary result set.



Continue with step 5.

3.2. If a @scope attribute is present:

3.2.1. If its value is "rscope:metadata": only metadata properties are in the scope of this rightsInfo element, add only all metadata elements of this item to a temporary result set. Continue with step 5.

3.2.2. If its value is "rscope:content": only content is in the scope of this rightsInfo element, add only all content of this item to a temporary result set. Continue with step 5.

4. If an @idrefs attribute is present in the base:

4.1. Iterate over each of the IDs listed by the @idrefs attribute and find the referenced element:

4.1.1. If the referenced element is a partMeta element then check if a @scope attribute is present in the base:

4.1.1.1. If a @scope attribute is not present: a) the partMeta content and b) all the partMeta metadata properties are governed by the base's rights expression; they all should be included into a temporary result set. Continue with step 5.

4.1.1.2. If a @scope attribute is present:

4.1.1.2.1. If its value is "rscope:metadata": only metadata properties are in the scope of this rightsInfo element, add only the metadata elements of this partMeta element to a temporary result set. Continue with step 5.

4.1.1.2.2. If its value is "rscope:content": only content is in the scope of this rightsInfo element, add only the content described by this partMeta element to a temporary result set. Continue with step 5.

4.1.2. If the referenced element is not a partMeta element: add the referenced element to a temporary result set. In this case the scope is implied by the element that is referenced and any @scope attribute should be ignored. Continue with step 5.

5. Check the base for an @aspect attribute:

5.1. If an @aspect attribute is not present then all members of the temporary result set should be copied to each of the result sets for the different Rights Aspects.

5.2. If an @aspect attribute is present then all members of the temporary result set should be copied only to the result sets corresponding to the Rights Aspects which are present in the @aspect list.

6. Final step: iterate over the result sets for the different Rights Aspects and interpret the included parts of the content or metadata elements according to the associated aspect. Some members of the result set may not be in a scope specified in the definition of the aspect; such members should be excluded from the result set.

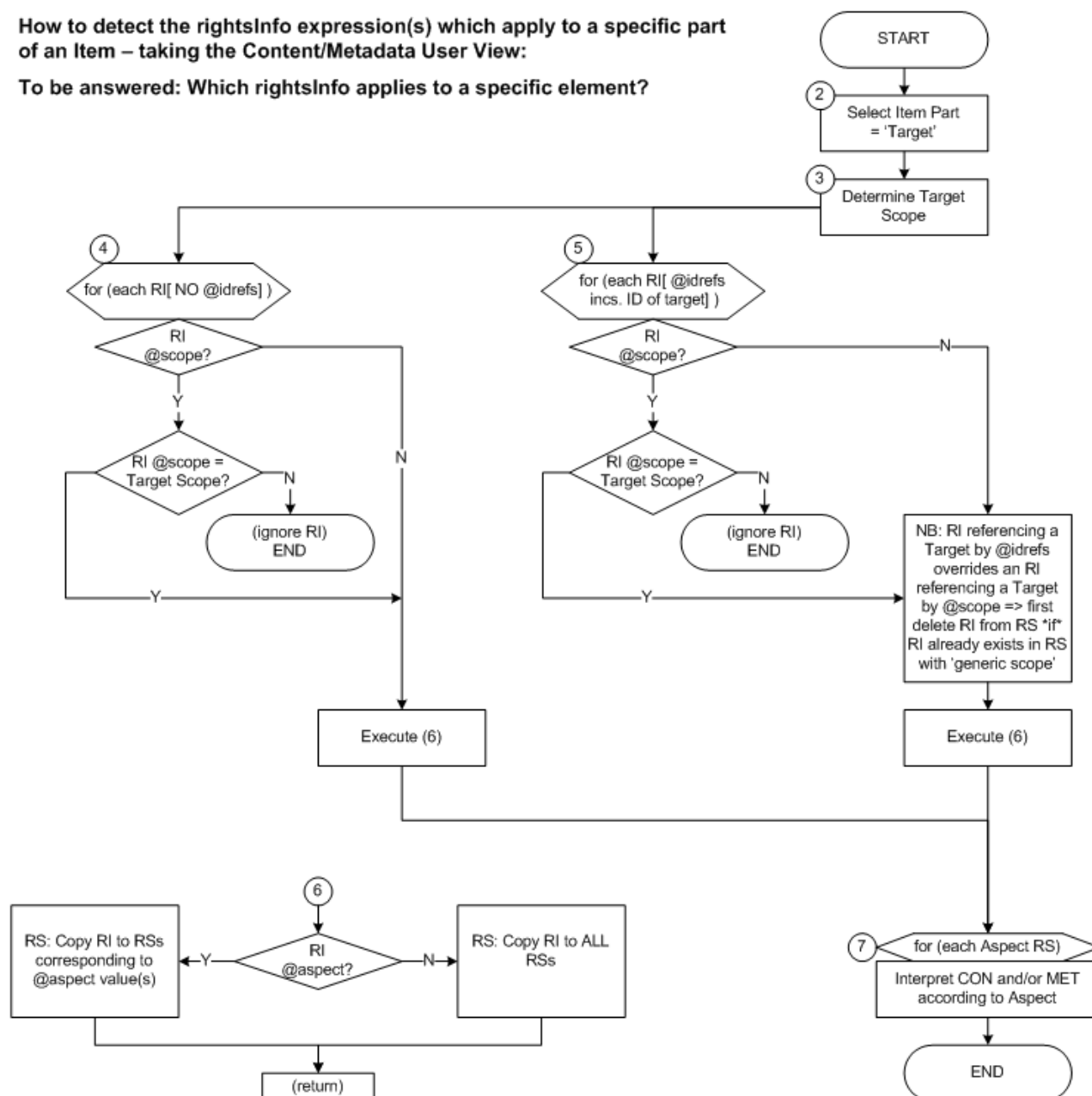
How to detect the rightsInfo expression(s) which apply to a specific part of an Item - taking the Content/Metadata User View:

To be answered: Which rightsInfo applies to a specific element?

Use Case 3 / 3:

How to detect the rightsInfo expression(s) which apply to a specific part of an Item – taking the Content/Metadata User View:

To be answered: Which rightsInfo applies to a specific element?



1. The goal of the processing: the result will be multiple sets of rightsInfo elements which all will apply to this part of the Item. Each of the sets a) correspond to one of the Rights Aspect NewsCodes, and b) may be empty after the processing if no corresponding rightsInfo elements were found.
2. Select the part of the Item for which the corresponding rightsInfo expression(s) should be determined, this part is the "target" for all subsequent processing steps.

This part must be a) the full content, b) one of the renditions of the content as a whole, c) a part of the content which is described by a partMeta element, or d) a single metadata property. The metadata wrappers itemMeta or contentMeta should not be selected as a target of this processing.

3. Define into which scope of rightsInfo elements the target falls:

Match the target against the definitions of corresponding parts for "riscope:content" and



"rscope:metadata" of the Rights Info Scope NewsCodes and determine to which scope the target belongs.

Be aware that partMeta elements fall under both scopes.

4. Iterate over each rightsInfo element which has no @idrefs attribute:

4.1. If a @scope attribute is not present in the rightsInfo element then check the rightsInfo element against the rules of step 6 and add it to result sets as defined. Earmark the added rightsInfo element as "generic scope rightsInfo". Continue with step 7.

4.2. If a @scope attribute is present and the target falls in the scope of the attribute's value (see step 3) then check the rightsInfo element against the rules of step 6 and add it to result sets as defined. Earmark the added rightsInfo element as "generic scope rightsInfo". Continue with step 7.

5. Iterate over each rightsInfo element which has an @idrefs attribute that includes the ID of the target:

5.1. If a @scope attribute is not present then check this rightsInfo element against the rules of step 6. Be aware that a rightsInfo element which is referencing the target by @idrefs overrules rightsInfo elements which reference the target by @scope. For that reason if the target should be added to the result set then first delete any rightsInfo element which is earmarked as "generic scope rightsInfo" from the result set, and then add this rightsInfo element. Continue with step 7.

5.2. If a @scope attribute is present and the target falls in the scope of the attribute's value (see step 3) then check the rightsInfo element against the rules of step 6. Be aware that a rightsInfo element which is referencing the target by @idrefs overrules rightsInfo elements which reference the target by @scope. For that reason if the target should be added to the result set then first delete any rightsInfo element which is earmarked as "generic scope rightsInfo" from the result set, and then add this rightsInfo element. Continue with step 7.

6. Check any @aspect attribute of a rightsInfo element:

6.1. If an @aspect attribute is not present then the rightsInfo element should be added to the result sets corresponding to each of the Rights Aspect NewsCodes.

6.2. If an @aspect attribute is present then the rightsInfo element should be added only to the result sets corresponding to the Rights Aspects which are present in the @aspect list.

7. Final step: iterate over the result sets for the different Rights Aspects and interpret the included parts of the content or metadata elements according to the associated aspect. Some members of the result set may not be in a scope specified in the definition of the aspect; such members should be excluded from the result set.

2.7 Item Metadata

Such information is wrapped in the mandatory **itemMeta** wrapper element and split between news management metadata and Item links.

2.7.1 Management Metadata

Management metadata is bound to the Item as a whole and reflects its processing in a professional workflow.

The order of the properties in this set is imposed by the W3C XML Schema.

Table 1. Item Management Group Elements

Element Title	Element Name	Card	Described on Page
Item Class	itemClass	(1)	201
Content Provider	provider	(1)	118
Date Item Version Created	versionCreated	(1)	133
Date Item First Created	firstCreated	(0..1)	132
Date Item Embargo Ends	embargoed	(0..1)	131
Publish Status	pubStatus	(0..1)	261
Role in the Workflow	role	(0..1)	279

File Name	filename	(0..1)	167
Generator Tool	generator	(0..1)	170
Profile	profile	(0..1)	259
Editorial Service	service	(0..unbounded)	154
Item Title	title {itemMeta}	(0..unbounded)	208
Editorial Note	edNote	(0..unbounded)	153
Member Of	memberOf	(0..unbounded)	219
Instance Of	instanceOf	(0..unbounded)	198
Signal	signal	(0..unbounded)	292
Alternative Representation	altRep	(0..unbounded)	75
Deliverable Of	deliverableOf	(0..1)	147
Hash Value	hash	(0..unbounded)	147

The IPTC provides a mandatory standardised scheme applicable to the **itemClass** property of a newsItem, identified by the URI <http://cv.iptc.org/newscodes/ninature/>.

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

2.7.2 Processing the Publish Status of an Item

The IPTC makes these values normative for the exchange of Items between a provider and its customers:

- ◆ Usable: The Item MAY be published without restriction.
- ◆ Withheld: Until further notice, the Item MUST NOT be published or used under any circumstances. If the Item has been published the publisher MUST take immediate action to withdraw or retract it.
- ◆ Canceled: The Item MUST NOT be published or used under any circumstances. If the Item has been published the publisher MUST take immediate action to withdraw or retract it.

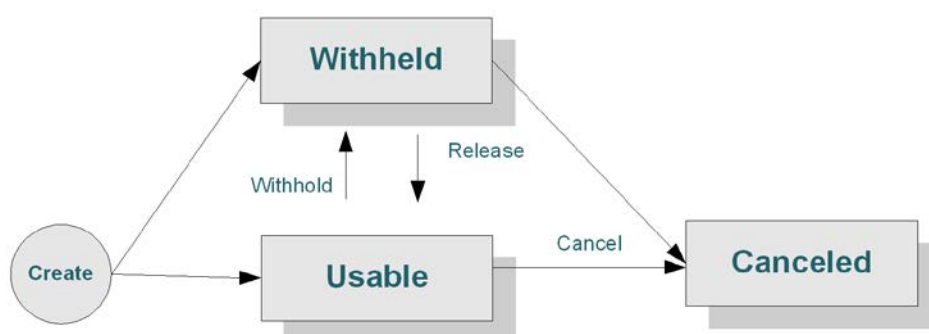
Embargoes are managed by the **embargoed** property. At the level of the data model the embargoed element could be linked now to an **edNote** element if the existing embargoed is empty (<embargoed />).

Details are described in the Processing model below

2.7.2.1 State Transition Diagram

This depicts the state transition diagram reflecting the ways in which the **pubStatus** values are intended to be used. Thus, upon creation of an Item, allowed statuses are usable and withheld. It is possible to withhold a usable document; it is possible to release a withheld document; it is possible to cancel a usable or withheld document. Once an Item has had its status set to canceled, it has reached a final state.

Figure 1. State Transition Diagram



2.7.2.2 Use Cases Associated with a Status of Withheld

Use Case 1: A provider distributes a story as a newsItem (version 1) with the status usable. At a later stage he learns that there may be a problem with the information included in the Item. He sends a new version of the newsItem (version 2) with a status set to withheld. All recipients systems must display a warning on this



newsItem, and recipient publishers must postpone the publication of the information contained in the newsItem until further notice. The news provider has confirmation that the information is false and decides to set the status to canceled (version 3).

Use Case 2: An eCommerce system proposes a large collection of illustrated articles managed as news items. The publisher managing the system sees that the information included in a newsItem (version 1) is not up to date anymore, and decides to hide this Item from its customers until it is properly revised. He set then its status to withheld (version 2), edits the newsItem and set its status back to usable (version 3).

2.7.2.3 Processing Model on the Recipient Side

Here is the processing model on the recipient side and relies on the **pubStatus** and embargoed properties:

Test pubStatus = canceled:

The Item must not be used, ever. Any usage of the Item must be prohibited, if needed by the way of alerts.

Else: next

Test pubStatus = withheld:

The Item must not be used until further notice. Any usage of the Item must be prohibited, if needed by the way of alerts.

Else: next

Test pubStatus = usable:

Test embargoed as described in [Table 2](#).

Table 2. Test pubStatus = Usable

<embargoed>	<pubStatus>	How to Process
Element is absent.	Usable	Item is usable and not embargoed.
Element exists, provides a Date/Time value.	Usable	The embargo on the item ends at the given date and time.
Element exists, but is empty. Corresponding edNote exists.	Usable	The item is embargoed as long as a condition applies which is described in an editorial note. The edNote should be formulated like this: <edNote @role="noteRole:embargo">Until end of speech</edNote>
Element exists, but is empty. No corresponding edNote exists.	Usable	The item is embargoed indefinitely. This may be overridden by a contractual agreement between the provider and the client.

2.7.3 Processing of versionCreated

If the value provided by any date/time field does not conform to the appropriate syntax (e.g. format "YYYY-MM-DDTHH:MM:SS[+-]HH:MM:SS") it MUST be considered as being not existent.

In the case of the mandatory **versionCreated** property the full Item MUST be considered as being void.

2.7.4 Best Practice for expressing an update or correction of an item

An Update is expressed by using the concept URI <http://cv.iptc.org/newscodes/signal/update> (as QCode with the recommended scheme alias: sig:update) as value of the Signal property under the Item Meta of an Item. This signal indicates that some part of the item has been updated. This implies that this version of the item is not the initial version.

A Correction is expressed by using the concept URI <http://cv.iptc.org/newscodes/signal/correction> (as QCode with the recommended scheme alias: sig:correction) as value of the Signal property under the Item Meta of an Item. This signal indicates that some part of the item has been corrected. This implies that this version of the item is not the initial version. This Correction signal does not indicate in which version(s) of the item the corrected error existed.



In addition a concept from the Severity NewsCodes (<http://cv.iptc.org/newscodes/severity/>) may be used as a refinement of how severe the impact of this update or change is. The IPTC acknowledges that the rules for applying the severity are set by the news provider of the item.

Further the Editorial Note (edNote) property under Item Meta may be used to provide details about the update or correction like pointing at a name in the text which has been corrected or if paragraph with updated information has been added to the text.

2.7.5 Best Practice for issuing a content warning

A Content Warning is expressed by using a QCode for the concept URI <http://cv.iptc.org/newscodes/signal/cwarn> with the **signal** property. (With the recommended alias the QCode is “sig:cwarn”.) This signal indicates that the content of the item should be reviewed as it may be perceived as being offensive.

In addition, refinement of the reason(s) for the content warning MAY be expressed by using concept(s) from the Content Warning NewsCodes <http://cv.iptc.org/newscodes/contentwarning/> with the **exclAudience** property.

Examples:

1. Content Warning signal without specific Content Warning NewsCodes:

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

2. Content Warning signal with specific Content Warning NewsCodes (relating to nudity and language):

```
<signal qcode="sig:cwarn"/>
<exclAudience qcode="cwarn:nudity"/>
<exclAudience qcode="cwarn:language"/>
```

14.3 2.8 Item Links

A powerful feature of NewsML-G2 is the capability to associate Items via links. It is therefore possible to create a network of news resources, for management and navigation purposes.

The **link** element offers a generic mechanism for linking Items within the NAR framework as well as creating links from Items to other Web resources.

The semantic of the link MAY be refined via a relationship attribute (**rel**). In the absence of such indicator, the implied meaning of the link is “see also” (i.e. a navigation link).

The IPTC provides a recommended scheme of link relationships identified by the URI <http://cv.iptc.org/newscodes/itemrelation/>.

To identify the target resource either the **residref** attribute or the href attribute MUST be set, optionally both MAY be used in parallel. The **residref** attribute identifies the target resource by its globally unique identifier (if the resource has such an identifier), while the **href** attribute identifies the location of the target resource in e.g. a (remote) file system. If the target resource is an Item and the **residref** attribute is used, a version attribute MAY indicate the target Item version; in the absence of version information, the target resource is the latest version available.

The content type, a.k.a. IANA MIME type of the target resource MAY be indicated by the contenttype attribute. It MAY be complemented by a format attribute to refine the MIME type information.

In order to ease the processing of a linked resource, the size in bytes of the target resource MAY be indicated. This feature is useful if the target on the link is a Web resource. If the target resource is an Item, the size which is given here MUST be the size of the XML representation of the Item.

A **rank** attribute may represent the rank of the link among other links.

This property also provides **timeValidityAttributes** (**validfrom** and **validto**) which express the date and time between which the link is valid.

Supplemental metadata extracted from the target resource (usually an Item) may be added to the linking information as child elements. Such information is not constrained by the data model. It may be part of the target Item Metadata (e.g. Publish Status, Alternative Location ...), Content Metadata (e.g. Intended Audience, Subject, Genre ...) or Characteristics of the content (e.g. Size, Content Type, Format, or specific



characteristics like the Height and Width of a picture). Different sets of characteristics may be provided, corresponding to specialized content components.

All properties SHOULD be included directly under the **link** property (see the details for this inclusion in the **link** specification table for the "Hint and Extension Point" child elements).

2.8.1 Processing Links

Link processing rules:

Link.1: Processor on the consumer side: If a **guid** and a **version** are provided, check whether the specific version of the Item is accessible using this information.

Processor on the provider side: If a **guid** and a **version** are provided deliver only the item version with the requested version number.

Link.2: Processor on the consumer side: If only a **guid** is available and no **version**, check whether an item is delivered by the provider. Consider a delivered version of the item as being the latest one.

Processor on the provider side: if only a **guid** is requested and not **version**, check if any version of the item exists, and if yes provide the one with the highest version number.

Link.3: Check whether the value of the **href** attribute allows some direct retrieval of the target resource via the Web (e.g. if the scheme is http: or ftp:), or an implicit resolution mechanism (e.g. DOI).

Link.4: Check whether an Alternative Representation (**altRep**) is exposed in the link. This information may complement the **href** attribute and provide an immediate URI resolution mechanism for Items. Multiple locations may be given, as allowed in the Item Metadata component. In such a case the processor will use the role qualifier and URL scheme for choosing the most appropriate resource.

Link.5: Signal an error or ignore the link.

2.9 News Content Metadata

News Content Metadata is directly associated with the news information conveyed by the Item, independently of the processing of the Item in a professional workflow. Such information which applies to the whole content of the Item is wrapped in the contentMeta wrapper element and split between administrative and descriptive metadata. Be aware that some G2 Items adopt only a subset of the metadata properties listed below. Information about a part of the content is wrapped by **Part of Content Metadata** (page 29).

2.9.1 Administrative Metadata

This is a set of properties associated with the administrative facet of content, i.e. data that cannot be inferred from "consuming" (reading, listening to, watching) the content.

All properties are optional. The order of the properties in this set is flexible: the non-repeatable properties MUST come first and then the repeatable properties may be inserted in any order.

Table 3. Administrative Metadata Group Elements

Element Title	Element Name	Card	Described on Page
Urgency	urgency	(0..1)	303
Date Content Created	contentCreated	(0..1)	129
Date Content Modified	contentModified	(0..1)	130
Located	located	(0..unbounded)	216
Information Source	infoSource	(0..unbounded)	200
Creator	creator	(0..unbounded)	125
Contributor	contributor	(0..unbounded)	120
Audience	audience	(0..unbounded)	80
Excluded Audience	exclAudience	(0..unbounded)	161
Alternative Identifier	altId	(0..unbounded)	73

2.9.1.1 Dates Processing Model



Two optional dates are associated with the content of an Item.

contentCreated and **contentModified** processing rules:

DatesGeneral.1: If the value provided by any date/time field does not conform to the appropriate syntax (e.g. format “YYYY-MM-DDTHH:MM:SS[+-]HH:MM:SS”) it MUST be considered as being not existent.

DateValues.1: If contentCreated is present it MUST NOT be later than **versionCreated** (page 133).

Error handling if it is later: at the creator's site an error alert should be issued, on the receiver's site it should be set to versionCreated.

DateValues.2: If contentModified is present contentCreated SHOULD be present as well. In this case contentModified MUST NOT be earlier than contentCreated.

Error handling if it is earlier: at the creator's site an error alert should be issued, on the receiver's site it should be set to contentCreated

DateValues.3: If contentModified is present it MUST NOT be later than versionCreated.

Error handling if it is later: at the creator's site an error alert should be issued, on the receiver's site it should be set to versionCreated.

DateProcessing.1: The recipient processor MUST first check if a contentModified element is present.

DateProcessing.2: If not it MUST check if a contentCreated element is present.

DateProcessing.3: If not it SHOULD assume that the content was created at the time indicated by versionCreated element in itemMeta.

2.9.1.2 Audience Processing Model

Audience processing may be used to form ad hoc groups of recipients for which the Item is particularly significant or to filter out some users from the list of intended recipients of an Item.

The audience is expressed as a set of “positive” values (**audience** (page 80)) and a set of “negative” values (**exclAudience**). The logic is to make the content easy to find to the audience identified by the positive values, but keep this content away from the audience identified by the negative values. An attribute of each property may indicate the expected significance of the content for this specific audience, and acts as a threshold for recipient filters.

The model for the audience processing is only a part of the overall filter that is used to determine whether a particular recipient is entitled to have access to the Item. It could be combined with the processing of other properties to further narrow the number of Items that match the recipient profile.

The processing rule has to be considered as a function which returns TRUE to indicate the recipient is entitled to receive the content, FALSE in case he is not entitled and NULL if the item does not contain any audience statements that apply to the Recipient.

Audience processing rules:

Audience.1: If any of the exclAudience properties applies to the recipient: return FALSE

Audience.2: If any of the audience properties applies to the recipient: return TRUE.

Audience.3: Return NULL.

2.9.2 Descriptive Metadata

This is a set of properties associated with the descriptive facet of news content, i.e. data that can be inferred from “consuming” (reading, listening to, watching) the news.

All properties are optional, repeatable and may be inserted in any order.

Table 4. Descriptive Metadata Group Elements

Element Title	Element Name	Card	Described on Page
Language	language	(0..unbounded)	211
Genre	genre	(0..unbounded)	171



Keyword	keyword	(0..unbounded)	209
Subject	subject	(0..unbounded)	296
Slugline	slugline	(0..unbounded)	293
Headline	headline	(0..unbounded)	182
Dateline	dateline	(0..unbounded)	143
By	by	(0..unbounded)	84
CreditLine	creditline	(0..unbounded)	126
Description	description	(0..unbounded)	146

2.9.3 Other Content Metadata

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

2.10 Part of Content Metadata

Streamed content may be split into different sections (called “shots” in the video world). Images may also be split in regions.

A specific set of metadata MAY be associated with any individual content part. Such metadata is wrapped in a **partMeta** element, which is repeatable in the newstern and MUST be inserted after contentMeta.

Each part MAY have a part identifier (partId) and a sequence number (seq).

Each part MAY be illustrated by an **icon** - e.g. a keyframe of a video clip - which takes the form of an IRI. It is not mandatory for such icon to be a pure extraction of the content. If multiple icon elements are present they MUST represent the same visual content, only differentiated by rendition, contentType or format.

A section of a stream MAY be defined by a **timeDelim** element. The time scope is expressed as **start** and **end** timestamp attributes plus an additional time unit (**timeunit**) attribute. Both timestamp values MUST be within the overall content duration.

A region of an image MAY be defined by a **regionDelim** element. Currently regions are limited to rectangles defined by {x, y, width, height} coordinates in pixels expressed as a set of attributes.

The role of this part in a stream of content MAY be defined by the **role** property.

If, during the processing of the content, it appears that part delimiters do not correspond to any physical content, then the corresponding set of metadata MUST simply be discarded.

News Administrative and Descriptive Metadata may be applied to each part, in complement to the administrative and descriptive metadata applicable to the whole content.

Each provider may add a set of metadata properties which have to be defined in a non-IPTC -G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

2.10.1 Edit Units and Time Codes

It is recommended that time and durations are expressed in “edit units” (EditUnit), which represent the smallest editable portion of content, i.e. a video frame or an audio sample.

$\text{EditUnit} = 1 / \text{EditRate}$.

For video, the EditRate is the FrameRate.

For audio, the EditRate is the SampleRate.

The use of EditUnit is independent of the mode of representation of time (e.g. timecode) in editing devices. The timecode associates one value to each video frame or audio sample.

For video, the usual timecode format is HH:MM:SS:FF (Hours:Minutes:Seconds:Frames).

In the case of simple frame rates (e.g. 25 fps, 30 fps, 50 fps or 60 fps), the conversion of a number of EditUnits to timecode is simple.

However, there exist other frame rates (e.g. 29.97 fps, 59.94fps) for which this calculation requires more attention. A precise calculation would consist of replacing e.g. 29.97 fps by its exact value $1.001/30$ fps and



multiplying the number of EditUnits by 1.001 before conversion on the basis of 30 fps. Another method consists of calculating the timecode using the drop frame method defined in SMPTE 12M. The drop frame method is an approximation based e.g. on 29.97 fps (1.001001001/30 fps). The drop frame timecode is not systematically used, particularly if content is of a short duration with insignificant drift with the actual clock time. SMPTE 12M will evolve as it doesn't address higher frame rates with progressive scanning.

For audio, the usual video timecode (HH:MM:SS:FF) is used if the content also contains video. A time restricted timecode (HH:MM:SS) is often used for audio only content.

The time reference will be the one of reception or edition in the production system, which should be able to locate content in time based on the number of EditUnits.

2.10.2 Time Unit Types and Start/End Timestamp Formats

The format of the Start Timestamp (@start) and/or End Timestamp (@end) is implied by the associated Time Unit type (@timeunit), see the Time Delimiter element [timeDelim](#).

Table 5 defines the processing of values of the three related attributes - but be aware: they are required by the XML Schema but may either show invalid values or be empty.

Table 5. Time Unit Type and Start/End Value Processing

Time Unit Type [@timeunit]	Start/End Timestamp [@start / @end]	How to Process
Invalid value	None	Ignore the Time Delimiter.
Invalid value	One or both	The default Time Unit Type value of editUnit MUST be used; the related format is used to parse the Timestamp value(s).
Valid value	None	Ignore the Time Delimiter.
Valid value	One or both	The defined Time Unit Type value MUST be used; the related format is used to parse the Timestamp value(s).

2.11 Assertions About Concepts

When a concept is used as the value of many properties or by a property with a limited granularity of concept details, it may be useful to group supplemental information about this concept at a unique location.

The optional and repeatable [assert](#) property provides information about a concept identified by a qualified code, a full URI or a literal value. The information is given as a set of properties providing metadata about the concept. Many assertions may be included in an Item.

Any property of the concept may be included at this point, especially its name, its relationships with other concepts, its definition.

Note: This information is only up to date at the time of last modification of the Item. Any changes applied to a concept after that time are not reflected by an assert element.

2.12 References to Inline Concepts

When the same concept appears as a string in several different labels or in the textual content of a newsItem, it may be useful to group information about this concept at a unique location.

The optional and repeatable [inlineRef](#) property provides information about a concept found in some textual content. The string associated with the concept can be tagged by any element which provides an attribute of type ID. One or more local identifiers MAY be listed as value of the *idrefs* attribute of the inlineRef element.

If the concept is taken from a controlled vocabulary it MUST be identified by a qualified code or a full URI, in any other case it SHOULD be identified by a literal value, and supplemental information MAY be given as a set of properties relative to the concept.

It is possible to give values for the confidence with which the metadata has been assigned, the relevance of the metadata to the string to which it is attached, and why the metadata has been included.



2.13 Document Derivation of Concepts

An increasing number of metadata values is not added explicitly by human interaction but by an automated derivation using some kind of knowledge network. In this case it could be of value to indicate the concept(s) from which a specific value of a metadata property has been derived. For this purpose the optional **derivedFrom** element can be used.

The qcode or uri attribute of this element define the concept from which another concept has been derived. The idref attribute of this element refers to the id attributes of all the properties in this G2 item having a value which has been derived from the concept represented by the derivedFrom property.

2.14 newItem Content

Content may be included by value or by reference, and useful characteristics are represented along with such content, in order to facilitate its processing.

Alternative renditions of the news content, i.e. different technical representation of the same logical content, are wrapped by a **contentSet** wrapper element. Their order of appearance in contentSet is of no relevance. Their presence is optional: this allows for a lightweight and extensible representation of information.

Each rendition SHOULD be defined by a **rendition** attribute.

All alternative renditions SHOULD be derived from an original rendition by a software processor. For example: images in different resolutions, vector graphics and alternative bitmap images, text in different formats (e.g. NITF and PDF). The rendition from which all other renditions originate is indicated by the **original** attribute of contentSet; this attribute takes as a value the local identifier (id) of the original content component included in the contentSet.

They are three kinds of content components, inline XML, inline data and remote content:

- ♦ The **inlineXML** wrapper element holds XML content which is directly embedded in the element. The root element of this structure must be the root element of the language. Content may belong to any XML language capable of expressing generic or specialized news information, e.g. NITF, XHTML, SportsML or XBRL. The XML vocabulary is identified by a content type attribute (**contenttype**).
- ♦ The **inlineData** wrapper element holds plain-text or base64 encoded content. Plain text or CDATA content MUST be identified by the "text/plain" content type. Binary content, like images, audio clips or even PDF or Word documents may be exchanged after proper encoding, but it is strongly recommended to use this structure for small assets only. The encoding algorithm MAY be indicated using the **encoding** attribute. In the absence of this attribute, the content must be plain text, and the content type must be set accordingly. Encoding is not constrained to base64 at this level of conformance.
- ♦ The **remoteContent** wrapper element may be used for representing any kind of media and data format. The data is stored independently of the newItem and is referenced via a hyperlink (href). The size in bytes of the remote content MAY be indicated. The element MAY also have **time-ValidityAttributes** (**validfrom** and **validto**) which express the date and time between which the reference is active. The same rendition of content MAY be present at different remote locations. In this case alternative locations of the content are provided by altLoc child elements of one remoteContent, multiple remoteContent elements with the same rendition value SHOULD NOT be used.

The description of the content in each content component MAY be complemented by a content type (contenttype), a format acting as an optional refinement of the content type, an indication on the software tool used to generate the content (generator) and the date and time when the content was generated, plus additional news content characteristics.

All these three types of content component elements have an id attribute. For this attribute a special constraint applies: its value MUST be persistent for all versions of the item, i.e. for its entire lifecycle. The rea-

son for this constraint is that G2 elements referencing a target G2 item may further point inside this item to reference a specific content component by its - persistent - id.

2.15 News Content Characteristics



newsContentCharacteristics are these physical properties of media content like the height and width of a picture, the word count of a news story or the duration of an audio clip, that help making selections among alternaterenditions of news content.

The characteristics defined by the IPTC are a small and typical set of properties. Individual providers may add more characteristics they consider reasonable, i.e. audio data for professional broadcasting may require a different set from audio content for a podcast.

2.16 Channels

Some binary streams support the notion of channel or track: this is e.g. the case for DVD's, which are MPEG-2 encoded and provide several audio tracks in different languages. It may be important to indicate media characteristics on a per-channel level.

A repeatable **channel {News Item}** element MAY therefore be defined as a child of a **remote- Content** element.

Each logical channel MAY have a local identifier (*chnid*), an indication of the media type of the data conveyed by the channel and an indication of the role the data plays in the scope of the full content, e.g. "voice over".

Each logical channel MAY be additionally described by the news content characteristics corresponding to the media conveyed in the channel.



3 Introduction to EventsML-G2

EventsML-G2 is a member of the Family of IPTC G2-Standards which is built on a common structural and function framework called the IPTC News Architecture (NAR). The EventsML-G2 specification extends the NewsML-G2 structural specification with some event-specific details and adds well defined functionality for conveying events.

3.1 Overview

3.1.1 What is EventsML-G2?

- ◆ EventsML-G2 is a standard for conveying event information in a news industry environment.
- ◆ EventsML-G2 is a member of the Family of IPTC G2-Standards; this family builds on a common specification for the exchange of news items and knowledge about topics, concepts and events.
- ◆ EventsML-G2 may be used for:
 - Receiving all facts about a specific event from the event organiser
 - Publishing all facts about a specific event by a news provider
 - Publishing all or only a subset of the facts of one to many events by event listings
 - Storing facts about knowledgeable events in archives to be referenced by other items

3.1.2 Business Advantages of Using EventsML-G2

EventsML-G2 is:

- ◆ Comprehensive (= many types of events may be covered).
- ◆ Flexible (= copies of substructures may be used many times, e.g. all the people appearing at an event).
- ◆ Extensible (= news provider specific data structures may be added to capture further facts about events)

EventsML-G2 may express facts and information about events by concepts identified either by literal text (free text) or by codes from controlled vocabularies.

EventsML-G2 provides very flexible date types:

- ◆ year, month, day, optionally plus time
- ◆ year and month only or even year only
- ◆ approximative dates = a date range

EventsML-G2 reuses building blocks from the common NewsML-G2 Architecture allowing for the reuse of software components, making their implementation cheaper.

EventsML-G2 makes use of industry standards: allows processing with standard tools. The EventsML-G2 syntax is built on XML, the Extensible Markup Language of the W3C, furthermore, EventsML-G2 makes use of W3C XML Schema and complies with the basic notion of the Semantic Web, the Resource Description Framework (RDF). This allows an easy transfer of EventsML-G2 structures to other XML-based standards and the integration of information about an event into the Semantic Web.

3.1.3 What is an Event – to be represented by EventsML-G2

An event is “something that happens” by definition. For the news industry, it is “something that happens and is subject to news coverage.” All the events in a day make up a “daybook”, which can be a marketable product sold to clients or simply an internal daybook used by editors to organise their work.

An event is planned or unplanned, with breaking news capable of overshadowing everything on the schedule.

Automated systems need to store and exchange information about news events. This is currently done in an ad-hoc manner, leading to overly-specialized formats and incompatible data exchange. From that the



IPTC learned that the industry would benefit from an event information interchange standard. Such a standard would facilitate the efficient exchange of event information, and the creation of better tools to support event management.

Note: information about the planned coverage of an event can be shared by using a Planning Item - see [Planning news coverage - planningItem](#)

3.2 Definitions

3.2.1 Event Information

The event information describes a particular event in detail. This includes the “who”, “what”, “when”, and “where” information for the event along with identification and publication (news management) information. The event information also includes facilities for relating events to each other and relating news items (both complete and incomplete) to the event information.

3.2.2 Coverage Information (LEGACY)

Note: the G2-Standards have a newer and more powerful tool for expressing and managing the planned coverage of events: [Planning news coverage - planningItem](#) . To provide backward compatibility the structure for coverage information as part of an event structure is still valid, but it is strongly recommended to separate out the planning information into the Planning Item, enabling event definition and planning to be decoupled.

The old-style coverage information describes the plan of news coverage for this event but it is highly recommended to adopt the new-style Planning Item.

3.2.3 The Data Model

The data model for EventsML-G2 has to cover two different facets of event information which relate to a basic distinction made for all G2 standards:

- ◆ Persistent Knowledge: is information which is remembered and referenced to for a long time.
- ◆ Topical News: is typically volatile information in the sense of “nothing is older than yesterday's news”.

For EventsML-G2 this is reflected by two different data models:

- ◆ Persistent information about an event is represented by an NewsML-G2-Concept Item which is a generic NAR structure for concepts extended by a set of detailed information specific to an event. As for any other kind of Concept also this specific one for events can be referenced by its Concept Identifier.

The same applies to Knowledge Items: a variant with event specific extensions is available, in particular event details are added to the concept structure inside the Knowledge Item. Knowledge Items may be used to exchange a set of event information if it should be distributed with a concept identifier.

Find details about this data model in section [An Event Concept in a Concept Item or Many Events in a Knowledge Item](#) .

- ◆ Volatile information about an event is represented by an “event” structure which is plugged into a NewsML-G2 News Item as its content. A single news item may include one to many event structures. This kind of event information cannot be referenced as persisting information from any other item. Find details about this data model in section [Events in a NewsItem](#) (page 39).

The most important thing to note about the EventsML-G2 data model is that the core structures holding information about an event are identical for both the content plugged into a News Item and the extension of a Concept Item. Hence it is very easy to build a single EventsML-G2 processor for topical and persisting information about an event.

3.3 EventsML-G2 and iCalendar

A well known and widely used standard for events data is “iCalendar” which is specified by RFC 2445.

EventsML-G2 compares very well to it as it covers virtually all features of an iCalendar Event Component:

**Table 6. iCalendar-to-EventsML-G2 Component Mapping**

iCalendar Event Component (Alphabetically)	Corresponding EventsML-G2 Component
attach	“link” property of a G2-item
attendee	“participant” property
categories	“subject” property
class	Access management functionality, no direct equivalence in EventsML-G2
comment	“note” property (under “event” for news and “concept” for a concept)
contact	“contactInfo” property (under eventDetails)
created	“contentCreated” property (in contentMeta) of a G2-item for news or a concept.
description	“definition” property (under “event” for news and “concept” for a concept)
dtend	“end” property (under eventDetails/dates)
dtstamp	“contentCreated” property (in contentMeta) of a G2-item for news or a concept.
dtstart	“start” property (under eventDetails/dates)
duration	“duration” property (under eventDetails/dates)
exdate	“exDate” property (under eventDetails/dates)
exrule	“exRule” property (under eventDetails/dates)
geo	“position” property (under eventDetails/location/geoAreaDetails)
last-mod	“contentModified” property (in contentMeta) of a G2-item for news or a concept.
location	“location” property (under eventDetails)
organizer	“organiser” property (under eventDetails)
priority	As this iCalendar property reflects the priority for a calendar of an individual no equivalent exists in EventsML-G2.
rdate	“rDate” property (under eventDetails/dates)
recurid	No direct equivalence in EventsML-G2, assigned functionality may be replicated by G2-item means.
related	No direct equivalence, but relationships can be expressed by other G2-item means
resources	Not covered by EventsML-G2 1.0, planned for future versions.
rrule	“rRule” property (under eventDetails/dates)
rstatus	Scheduling protocol functionality is not covered by EventsML-G2
seq	“version” attribute of the G2-item's root element

**Table 6. iCalendar-to-EventsML-G2 Component Mapping (Continued)**

iCalendar Event Component (Alphabetically)	Corresponding EventsML-G2 Component
status	“confirmation” (under eventDetails/dates) reflects the status of confirmation of the dates of the event, while “occurStatus” (under eventDetails) reflects the overall status of the event.
summary	“name” property (under “event” for news and “concept” for a concept)
transp	Not covered by EventsML-G2
uid	“guid” attribute of the G2-item's root element
url	No direct equivalence. For G2-items it may be defined individually by each news provider how to resolve the identifier of an G2-item to an accessible location.
x-prop	EventsML-G2 provides “Extension points” for this purpose.



4 Events

4.1 The Core Information about Events

Regardless of whether the information about an event is topical or persistent (see [The Data Model](#) on page 34) the same structure is used to mark it up.

The information about an event includes a set of more generic properties:

- ◆ A natural language **name** for the event. This name should be rather concise and can be expressed in different languages.
- ◆ A natural language **definition** for the event which can be more extensive than the name; it can explain facets in detail.

The **role** attribute of a definition could be used to provide variants of an explanation, e.g. a short one for overviews and a rather extensive one for a detailed presentation.

- ◆ A natural language **note** about the event. This could be an explanation of details or background information regarding the definition. Again this note can be expressed in different languages and can be qualified by a **role** attribute.
- ◆ The properties **sameAs {Relationship}**, **broader**, **narrower** and **related** can be used to define a relationship between this event and another event or concept.

In particular broader may be used to express that this event is a sub-event to another event, e.g. a break-out session of a big conference, one competition of the Olympic Games or one of the concerts of a festival.

A **related** property may be used to further qualify the nature of the event. Related can take either an arbitrary literal value or a value from a controlled vocabulary and could be used to express e.g. that this event is a concert, a hockey game or a press conference.

Additionally, a set of rather event-specific properties - wrapped by the **eventDetails** property:

- ◆ A **dates** sub-structure expresses the start date and the end date or duration of the event. This includes using the "approximative dates", i.e. a range of dates, this range as a kind of best guess or most likely date.
If this event is recurring this can be expressed by means of recurrence properties which align to equivalent properties of the iCalendar standard RFC 2445 (see more below).
- ◆ **occurStatus** indicates the status of the occurrence - if this is an unplanned or planned event, and if it is planned - how likely it is to occur.
- ◆ A set of **registration** information which defines how persons have to register for the event, this may include the accreditation of journalists.
- ◆ A set of **accessStatus** information.
- ◆ A set of **participationRequirement** properties. This can be used e.g. for expressing age limits - think of required parental guidance for movies - or for formal requirements for training course events.
- ◆ A set of **subject** properties expressing what the event is about. Be aware of the difference between a related and a subject property: related should indicate the nature of the event, what the event is, while a subject indicates applicable categories for what the event is about. For example, "concert" is a related concept, while "music" or "Wolfgang Amadeus Mozart" is a matching subject.
- ◆ A set of **location** properties. In most cases it will be the single location where the event will take place - but e.g. festivals could have more than one location.
- ◆ A set of **participant** properties to list all kinds of parties appearing in different roles at the event - the particular role can be expressed by the **role** attribute.
- ◆ A set of **organiser** properties to list all parties involved in organising the event - the particular role can be expressed again by the **role** attribute

- ◆ A set of **contactInfo** properties for the event. Be aware that the location, the participant and the organiser properties may contain contactInfo structures, but they pertain only to this particular property while this contactInfo is to be used for the event as a whole.
- ◆ A set of **language** properties reflecting all “official” languages at the event.
- ◆ A **newsCoverage {Concept}** property is still present in the specifications, purely for backwards compatibility; be aware that its status has changed to DEPRECATED in EventsML-G2 1.6. Conveying information about the planned coverage of an event should now use the generic **Planning news coverage - planningItem**.
- ◆ As for many wrapping elements in G2-Standards, the information about an event can also be extended by provider-specific properties.

4.2 Event Information in Items

4.2.1 An Event Concept in a Concept Item or Many Events in a Knowledge Item

The persisting knowledge facet of event information is represented as a Concept (see **Representing Concept Information - concept Component**)

As all other concepts a single event can be managed by a Concept Item (see **Managing Individual Concepts - conceptItem**), and subsequently many of them by a Knowledge Item (see **Managing Sets of Concepts - knowledgeItem**).

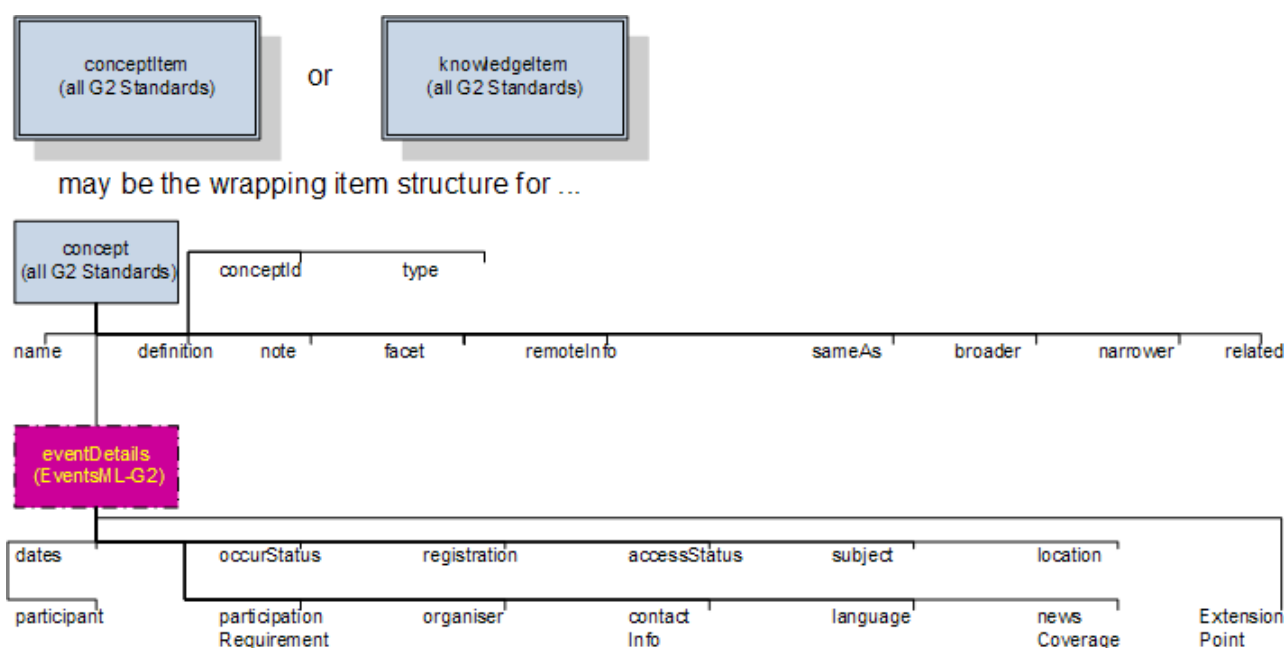
Any Concept Item or Knowledge Item provides a group of generic definitions and a set of details specific to a kind of concept, in this case specific to an event.

Event concepts use the generic part of a concept in order to define:

- ◆ The Concept Identifier for this event.
- ◆ A name, a definition, explanatory notes and refining related concepts.
- ◆ Relationships to other events.

In Event Concept Items the value of the type of a concept (conceptItem/concept/type) must be set to the concept URI of <http://cv.iptc.org/newscodes/cpnature/event> which may translate to a QCode of cpnat:event.

Figure 2. Event Information in a concept element



The event specific details are expressed by an eventDetails structure plugged into the “concept” of a Concept Item or a Knowledge Item. The eventDetails used there are completely identical to the structure with

the same name used for the “event” element in the content set of a News Item.

The Concept Identifier of an event can be used by other items (e.g. News Items or Concept Items) to reference this event. By a purely technical view this Concept Identifier can be used as value of any property referring to a concept. On a semantic level it is required that the semantics of this property includes events as allowed concepts – e.g. a property which is not limited to persons or locations by its semantics.

Examples are:

- ◆ Using an event's Concept Identifier as QCode for the “subject” property of a News Item. This indicates that the content of the News Item is about this event, the News Item's content may be text, photo, audio or video covering the event.
- ◆ Using an event's Concept Identifier with the **Same As**, **Broader**, **Narrower** and **Related Concept** properties of another Concept Item. By these means a structure or network of events can be created, e.g. to link individual performances with a cultural festival or different talks to a conference.

Knowledge Items with event concepts should be used to distribute event information if this information is planned to be updated - as this requires an identifier for each event.

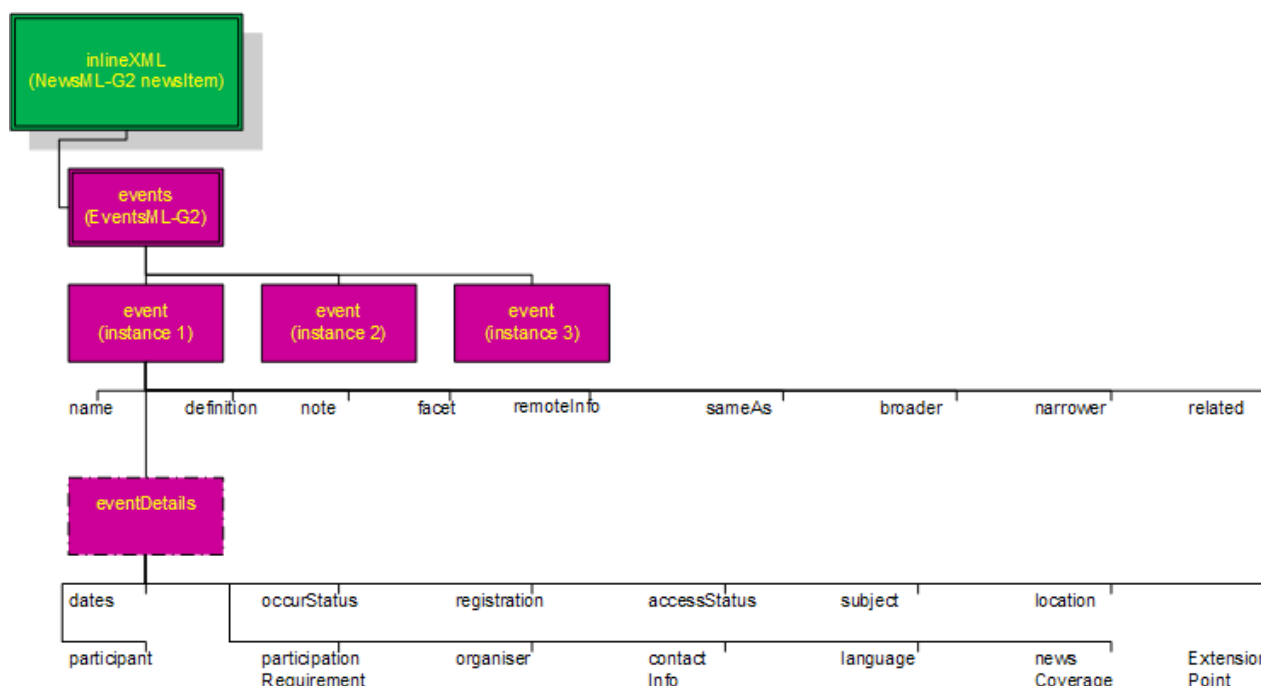
A provider could think of this use case scenario: a "top events of the next weekend" Knowledge Item is circulated with event concepts on Monday. On Wednesday, a new version of this Knowledge Item is sent with updated events, and cancelled events removed.

4.2.2 Events in a NewsItem

Topical event information may be conveyed by using the NewsML-G2 NewsItem (see [Representing News - newsItem](#)) as a wrapping item instance. The structure of a NewsItem defines a special node where content plug-ins can be attached: the inlineXML element.

For EventsML-G2 an **Events Wrapper** element has to be added as a child to **Inline XML** and it acts as a wrapper of one-to-many **Event** elements, each representing the topical information of a single event.

Figure 3. Event Information in a News Item



The event element wraps a group of more generic descriptions and a couple of details about an event. The first group is made of a short name which can be displayed as a one-liner, a more comprehensive definition of the event and a note with supplemental information.



A sibling to this generic group is `eventsDetails`, it wraps all the details of the event, when and where it happens, who is involved and how to get there.

Finally optional information about the planned news coverage of this item may be added.

4.2.2.1 News Metadata

In general the News Metadata section of a `NewsItem` - wrapped by the `Content Metadata {News Item}` element - should be populated and used as specified for NewsML-G2.

Further to this general recommendation these event specific considerations apply:

- ◆ If more than a single event is conveyed by a `NewsItem` the content metadata apply to the set of events as a whole. In most cases this set will be selected from a larger repository by some rules, like “events of next week”, or “music events”. This could be reflected by e.g. the headline, the description or even the subject property.
- ◆ Genre property: an appropriate value should be applied, like “almanac” or “daybook” from the IPTC Genre NewsCodes
- ◆ Language property: be aware of the difference between the language property of the content metadata - it reflects the languages used in the content, in this case in the description of the events - and the language property of the event structure - it reflects a language which is used at an event.

5 Representing Concept Information - concept Component

5.1 Concept Component

Concepts fall in two broad categories: named entities and generic (or abstract) concepts. Generic concepts range from themes (e.g. politics, soccer) to emotions (e.g. smiling, love); they have no specific property defined, beyond generic properties. Named entities are people, organisations, geopolitical areas, points of interest and objects for which a specific set of properties is defined for the purpose of a refined definition and improved search and processing capabilities.

The **concept** element provides a set of properties shared by all types of concept.

A concept can be identified in different schemes by different controlled values, this is why a concept identifier must be unambiguous, but cannot be unique: for example, a company may be identified by different identifiers from different company vocabularies. In the case of abstract topics, the strict sameness of two concepts may be subject to discussion, and therefore a notion of equivalence of concepts is preferred.

The properties common to all types of concepts are:

A concept **MUST** have a concept identifier, expressed as a **conceptId** child element.

The conceptId element **MUST** have a **qcode** attribute. It **MAY** have a **created** attribute and a **retired** attribute which limit the usage of the concept identifier in time.

A concept **MAY** have a **type** child element. The type of a concept reflects its nature, e.g. generic, person, organisation, geopolitical area, point of interest etc...

A concept **MUST** have a **name** and **MAY** be further defined in natural-language by a **definition** or **note** and by **remoteInfo**. Definition and note are repeatable and **MAY** be specified in multiple languages.

Different variants of a name are allowed. The **role** attribute refines the semantics of the property and takes values like "usual", "official", "married" (for a person) "acronym" (for an organisation), "synonym", "adjectival" (e.g. French for France). The **part** attribute identifies the part of the name conveyed by the property, and takes values like "given", "family" (for a person). Definitions and notes also support a role, which takes values like "history", "change" (for a description), "editorial", "scope" (for a note).

The descriptive elements definition, note and remoteInfo **MAY** have **validfrom** and **validto** attributes which limit the use of the property in time.

The remoteInfo element **MAY** be used to express any external information about the concept as such. Be aware that the link element in the itemMeta wrapper should only be used for linking a Concept Item as a whole to another resource, e.g. a previous version, or another ConceptItem from which this one was derived and not to resources relevant to describing the Concept.

A **hierarchyInfo** element **MAY** be used to express the location of this concept in the hierarchical tree of a taxonomy. For this purpose the hierarchyInfo holds a space separated sequence of the Concept Identifier QCodes of the ancestors of this concept, plus the Concept Identifier QCode of this concept. The sequence runs from left to right, with the top level QCode on the left, and the QCode of this concept on the right.

If the same concept is also defined in a different scheme this alternative identifier **MAY** be expressed by a **sameAs {Relationship}** child element.

The sameAs element **MUST** have either a qcode or a uri or a literal attribute which identifies a concept, for the exact rules see the table below in the chapter **Relationships Between Concepts**. It **MAY** additionally have a type attribute which reflects the nature of the associated concept, and **MAY** have one or more names (see **Flexible 1 Property Type**. **validfrom** and **validto** attributes **MAY** limit the relationship in time.)

More detailed properties of a concept (e.g. that the concept "is" an artist, listed company, city, restaurant) **MAY** be expressed by a specific **related** property. The related property **SHOULD** have a rel attribute which specifies the exact relationship between this concept and the target concept (e.g. "is a", "has a", "works for", "owns" ...). The IPTC provides a set of Concept Relationship NewsCodes for this purpose which is available at <http://cv.iptc.org/newscodes/conceptrelation/>.

5.2 Relationships Between Concepts

For any concept a relationship to another concept MAY be established, this may take form of a taxonomy (i.e. a hierarchy of concepts) or thesaurus (i.e. a set of concepts associated via standard relationships). A concept MAY establish a set of the most standard relationships **broader**, and **narrower** (page 222) and further MAY add a more flexible **related** relationship.

As the properties **sameAs**, **broader**, **narrower** and **related** establish a relationship to another property it is required to identify or describe this related concept. A specific selection out of three attributes **MUST** be used for this purpose. The basic rule is that all of them or none of them **MUST NOT** be used in any case. The following table defines how the attributes **MUST** be used with the different properties, when establishing a relationship. (Be aware that establishing a relationship to an arbitrary value is specific to the **related** property only, the specification section of this element provides more details.)

Table 7. Which attributes to use with relationship properties

Property	Attribute qcode or uri	Attribute literal	Set of attributes of an arbitrary value	Use case
sameAs	Yes	No	No	1
narrower	Yes	No	No	1
	No	Yes	No	2
broader	Yes	No	No	1
	No	Yes	No	2
related	Yes	No	No	1
	No	Yes	No	2
	No	No	Yes	3

Use cases for using the attributes to express the value to which the relationship should be established:

- 1) The value is a concept from a controlled vocabulary
- 2) The value is a concept which is not from a controlled vocabulary
- 3) The value is not a concept.

Further the **sameAs**, **broader**, **narrower** and **related** properties MAY have a type attribute which reflects the nature of the associated concept, and MAY have one or more names.

The **broader**, **narrower** and **related** properties MAY also have **validfrom** and **validto** attributes which limit the relationship in time, a **rel** attribute which details the name given to the relationship and a **rank** attribute which specifies the rank of the current concept among concepts having a relationship to the target concept. They also have a **facet** child property for expressing an intrinsic property of the related concept.

The **related** property has a **bag** property for allowing the expression of composite concepts (see **Composite Concepts**).

5.3 Details Associated with Specific Entities

Details associated with specific entities MAY additionally be defined. All have been chosen for their potential usefulness in the news industry.

personDetails include a date of birth (born) and date of death (died) a repeatable indication of affiliation with an organisation and contact information (contactInfo).

organisationDetails include a date of foundation (founded) and date of dissolution (dissolved), a repeatable location and contact information (contactInfo).

The registered address of an organisation is indicated as part of its contact information; if this address is used only for a formal registration and the organisations business office does not reside there it should not



be used for making direct contact with this company.

geoAreaDetails include the geographic coordinates (position) of the place.

The position **MUST** have *latitude* and *longitude* attributes. It **MAY** have an indication of the altitude above the zero elevation reference level.

It **MAY** have an indication of the coordinate reference system (*gpsdatum* attribute) expressed as a string. In the absence of this attribute, the WGS84 system is assumed.

POIDetails include the geographic coordinates (position) and the postal address of the place, plus practical information like opening hours (openHours), capacity, access information, details on the location (e.g. room number, stair number etc.), and contact information (contactInfo).

objectDetails include a date of creation, a creator of the object and a copyright notice.

5.3.1 Contact Information

contactInfo is repeatable in the definition of a person, an organisation and a Point of Interest, and each set of properties supports a *role* attribute which makes it possible to group together all information belonging of the same nature.

Contact information includes email addresses, instant messaging addresses (im), international phone numbers, international fax numbers, web addresses, postal addresses and notes. These are qualified by a *role* attribute which specifies the nature of the address, e.g. home or work.

5.3.2 Postal Address

The definition of a Postal Address (**address**) includes free-text lines (in the format expected by a recipient postal service), the indication of a locality (i.e. city, town, village etc...), a subdivision of a country (area), a country and a postal code (postalCode).

A postal address is structured as a set of properties likely edited and displayed as a form. The relative order of its properties is not universal, and if used for traditional postal mail, presentation algorithms are to be developed which depend on the source and recipient countries.

The city, country area and country may be indicated as a name or as a controlled value. The use of an ISO compliant country code is particularly recommended.



6 Managing Individual Concepts - conceptItem

An XML Schema file corresponding to the specifications for this item is available (see [The Full Set of Specification Documents](#) on page 3).

6.1 Description

A **conceptItem** aims to convey knowledge about a single concept, a named entity such as an organisation or an abstract notion such as a news subject (see [Representing Concept Information - concept Component](#)). Typically a conceptItem holds only a rather limited set of metadata about the concept and the structured concept data as content of the item.

Typical characteristics of a conceptItem are:

- ◆ It focuses on a single concept or entity.
- ◆ It will usually be updated infrequently but over a long period of time, when the information about the concept evolves.
- ◆ Its content is of long term interest.
- ◆ It may be referenced by other items.

Different Concept Items, managed by different providers, may contain structured information about the same concept.

6.2 Structure of a Concept Item

The model of a conceptItem is very similar to the model of a newsItem. Both share the same indicators of compliance with a standard and conformance level, Identification and versioning, Signature, Rights Information, Item Metadata, Item links. Please review the corresponding specification of a newsItem for more information.

6.2.1 Note about the different identifiers for a concept and a conceptItem

Each concept has its globally unique concept identifier (see [conceptId](#)) which is part of the concept structure and defined by the authority of the scheme.

Each conceptItem has its globally unique identifier (guid) attribute which is assigned by a system managing G2 items.

Be aware that these two identifiers must not be mixed up, all references to a concept MUST use the concept identifier and not the guid of the conceptItem.

6.3 Item Class

The IPTC provides a mandatory standardised scheme applicable to the **itemClass** property, identified by the URI: <http://cv.iptc.org/newscodes/cinature/>.

6.4 Concept related Metadata

The set of administrative metadata is common to all classes of Items.

The set of descriptive metadata for a Concept Item is listed below. All properties are optional, repeatable and may be inserted in any order.

Table 8. Descriptive Metadata Core Group Elements

Element Title	Element Name	Card	Described on Page
Language	language	(0..unbounded)	211
Keyword	keyword	(0..unbounded)	209
Subject	subject	(0..unbounded)	296
Slugline	slugline	(0..unbounded)	293
Headline	headline	(0..unbounded)	182
Description	description	(0..unbounded)	146

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2



namespace. See also [XML Namespaces](#) and [Extension Points in XML](#) .

Please review [News Content Metadata](#) of the News Item chapter for more information.

6.5 Metadata Helpers

The `conceptItem` includes three properties which are available to help make metadata assertions:

- ◆ the `assert` property: helps to bundle and extend details of concepts, see [Assertions About Concepts bookmark33](#)
- ◆ the `inlineRef` property: helps to reference concepts which are inline, within free text properties of type label, see [References to Inline Concepts](#) (page 30)
- ◆ the `derivedFrom` property: helps to document from which concept a concept, used as property value in this item, has been derived, see [Document Derivation of Concepts](#) (page 31)

6.6 `conceptItem` Content

The content of a `conceptItem` is a concept component (see [Concept Component](#) on page 41).



7 Managing Sets of Concepts - knowledgeltem

An XML Schema file corresponding to the specifications for this item is available (see [The Full Set of Specification Documents](#) on page 3).

7.1 Description

A **knowledgeltem** bundles a set of concept components which are managed and exchanged as a whole. A knowledgeltem is used best where a provider wants to circulate a snapshot of a set of entries from one or more controlled vocabularies.

The concepts represented in a knowledgeltem can be of different types, and their identifiers may come from different schemes. A “scheme definition” is therefore a particular case of structure, where all concepts support a concept identifier from a same specific scheme.

Examples of knowledgeltems are the taxonomy of IPTC Subject NewsCodes or an authority list of people's descriptions maintained by a given provider. Typical characteristics of a knowledgeltem are:

- ◆ It contains a set of concepts components covering a specific purpose, e.g. concepts from a single scheme, concept from different schemes and relevant in the context of a specific topic.
- ◆ It will usually be updated infrequently but over a long period of time, for example when a controlled vocabulary evolves.
- ◆ Its content is of long term interest.

7.2 Structure of a Knowledge Item

The model of a knowledgeltem is very similar to the model of a newsItem. Both share the same indicators of compliance with a standard and conformance level, Identification and Versioning, Signature, Rights Information, Item Metadata, Item links. Please review [Representing News - newsItem](#) for more information.

7.3 Item Class

The IPTC provides a mandatory standardised scheme applicable to the **itemClass** property, identified by the URI <http://cv.iptc.org/newscodes/cinature/>.

7.4 Knowledge Related Metadata

Metadata about the whole set of concepts held by a Knowledge Item are wrapped by the contentMeta element.

Metadata about specific concepts held by a Knowledge Item are wrapped by one to many partMeta elements. A typical use case of partMeta is to indicate when several concepts were modified at the same time, by associating those concepts with a specific partMeta which has the associated contentModified property.

The set of administrative metadata is common to all classes of Items.

The set of descriptive metadata for a Knowledge Item is listed below. All properties are optional, repeatable and may be inserted in any order.

Table 9. Descriptive Metadata Core Group Elements

Element Title	Element Name	Card	Described on Page
Language	language	(0..unbounded)	211
Keyword	keyword	(0..unbounded)	209
Subject	subject	(0..unbounded)	296
Slugline	slugline	(0..unbounded)	293
Headline	headline	(0..unbounded)	182
Description	description	(0..unbounded)	146

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

Please review [News Content Metadata](#) of the News Item chapter for more information.



7.5 Metadata Helpers

The `knowledgeItem` includes three properties which are available to help make metadata assertions:

- ♦ the `assert` property: helps to bundle and extend details of concepts, see [Assertions About Concepts bookmark33](#)
- ♦ the `inlineRef` property: helps to reference concepts which are inline, within free text properties of type label, see [References to Inline Concepts](#) (page 30)
- ♦ the `derivedFrom` property: helps to document from which concept a concept, used as property value in this item, has been derived, see [Document Derivation of Concepts](#) (page 31)

7.6 `knowledgeItem` Content

A `conceptSet` wrapper element contains a set of concept components (see [Concept Component](#)). Their order of appearance in `conceptSet` is not relevant.

Note: All concept definitions share the same catalog of schemes, declared at the top of the `knowledgeItem`.



8 Packaging Items - `packageItem`

An XML Schema file corresponding to the specifications for this item is available (see [The Full Set of Specification Documents](#) on page 3).

A `packageItem` facilitates the packaging of all kinds of Items, from really simple constructs to the highly hierarchical structures created by some news providers.

Examples of `packageItems` are a collection of pictures, a “top ten” list of `newsItems`, an unordered set of `newsItems` relative to the same event, the representation of a newspaper section or page.

Typical characteristics of a Package Item are:

- ◆ It provides some structure to news related information, and is expressed via a hierarchy of items.
- ◆ The Items found in a `packageItem` stay independent from the package: they can be managed individually, and the package keeps only references to them.
- ◆ Its content is of medium term interest.

8.1 Structure of a Package Item

The model of a `packageItem` is very similar to the model of a `newsItem`. Both share the same indicators of compliance with a standard and conformance level, Identification and versioning, signature, rights information, Item metadata, Item links. Please review the corresponding specification of a `newsItem` for more information.

8.2 Item Class

The IPTC provides mandatory standardised schemes applicable to the `itemClass` property of a `packageItem`, identified by the URI <http://cv.iptc.org/newscodes/ninature/> and <http://cv.iptc.org/newscodes/cinature/>.

8.3 Package Related Metadata

The set of administrative and descriptive metadata is common between `packageItems` and `newsItems`. Please review [News Content Metadata](#) of the News Item chapter for more information.

8.4 Metadata Helpers

The `packageItem` includes three properties which are available to help make metadata assertions:

- ◆ the `assert` property: helps to bundle and extend details of concepts, see [Assertions About Concepts bookmark33](#)
- ◆ the `inlineRef` property: helps to reference concepts which are inline, within free text properties of type label, see [References to Inline Concepts](#) (page 30)
- ◆ the `derivedFrom` property: helps to document from which concept a concept, used as property value in this item, has been derived, see [Document Derivation of Concepts](#) (page 31)

8.5 `packageItem` Content

A `groupSet` represents a tree of components, a component can be

- ◆ a `group` element which contains one to many of the components below
- ◆ an `itemRef` element referring to a package-external G2 item or a web resource
- ◆ a `groupRef` element referring to another group of this Package Item

All G2 items included into a package are included by reference, as physical inclusion would break the capability to manage inner Items independently of the outer Package Item.

The `groupSet` is optional. This allows for a lightweight and progressive representation of information.

There **MUST** be at least one group element in the `groupSet` but there could also be many of them. In any case the value of *root* attribute of the `groupSet` element **MUST** be the id attribute value of the group acting as a root.

A group component may contain references to other group components (using a `groupRef` element with its *idref* attribute) of the same package item and/or references to Items or Web resource (using the `itemRef`



element with its **guidref** and **href** attributes), in any order.

Each group **MUST** have an **id** attribute which identifies this group, and each group **MUST** have a **role** attribute which indicates the part this group plays within its container.

The order of the sub-groups and references to Items **MAY** be significant; a **mode** attribute indicates whether the elements in the group are complementary and unordered, complementary and ordered or a set of alternative elements. In the absence of a **mode** attribute the group is treated as complementary and unordered implementing the mode "bag".

The itemRef element **MAY** contain metadata extracted from the target Item or Web resource. The recipient **MUST NOT** consider that such hints constitute a complete representation of the Item.

The itemRef element **MAY** have a **rank** attribute which represents the rank of the Item among other Items in each group.

The itemRef element **MAY** also have time validity attributes (**validfrom** and **validto**) which express the date and time between which the reference is active.

Further there are other attributes available, see the specification table of the element.

Sample:

```
<groupSet root="g1">
  <group id="g1" mode="mode:seq" role="grouprole:main">
    <groupRef idref="g2"/>
    <itemRef guidref="urn:newsml:iptc.org:20070530:tutorial-text-xhtml"/>
  </group>
  <group id="g2" role="grouprole:gallery">
    <itemRef guidref="urn:newsml:iptc.org:20070530:tutorial-iptc-logo"/>
    <itemRef guidref="urn:newsml:iptc.org:20070530:tutorial-video"/>
  </group>
</groupSet>
```



9 Planning news coverage - planningItem

An XML Schema file corresponding to the specifications for this item is available (see [The Full Set of Specification Documents](#) on page 3).

9.1 Description

The **planningItem** facilitates conveying the planning of news and topic coverage from the editorial of the news provider to the editorials of his clients. This item was introduced with the EventsML-G2 1.6 and NewsML-G2 2.7 (both based on the News Architecture version 1.8) and it is intended to replace the information about planned newscoverage provided as sub-structure of the Event Details of an Event Concept Item by EventsML-G2. As the Planning Items is part of the common News Architecture framework it can be used in the scope of EventsML-G2 and NewsML-G2.

Typical characteristics of a planningItem are:

- ◆ It focuses on planning and delivering the coverage of a single event or topic but may be linked to other planning items to facilitate the coverage of e.g. large or long-lasting events or a group of topics.
- ◆ It will usually be updated frequently until all planned coverage is delivered
- ◆ Its content is a structured representation of typical parameters of editorial planning and further may provide a list of G2 items which have been delivered to fulfil the intended coverage.
- ◆ It may refer to the event it covers: examples are media events like press conferences, political events like an election, cultural events like an open-air concert, or sport events.
- ◆ It may refer to the topic(s) it covers: examples are topics like "The current housing market", "The cultural festival summer season in Europe", "The best skiing resorts in the Rocky Mountains".

9.2 Structure of Planning Item

The model of a planning item is very similar to the other G2 items: It shares the indicators of compliance with a standard and a conformance level, Identification and Versioning, Signature, Rights Information, Item Metadata and Item links. Please review [Representing News - newsItem](#) for more information.

9.3 Item Class

The IPTC provides a mandatory standardised scheme applicable to the itemClass property of a planningItem, identified by the URI <http://cv.iptc.org/newscodes/plnature/>.

9.4 Planning Related Metadata

The set of administrative metadata is common to all classes of Items.

The set of descriptive metadata is listed below. All properties are optional, repeatable and may be inserted in any order.

Table 10. Descriptive Metadata Core Group Elements

Element Title	Element Name	Card	Described on Page
Language	language	(0..unbounded)	211
Keyword	keyword	(0..unbounded)	209
Subject	subject	(0..unbounded)	296
Slugline	slugline	(0..unbounded)	293
Headline	headline	(0..unbounded)	182
Description	description	(0..unbounded)	146

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

Please review [News Content Metadata](#) of the News Item chapter for more information.

9.5 Metadata Helpers

The planningItem includes three properties which are available to help make metadata assertions:



- ◆ the assert property: helps to bundle and extend details of concepts, see [Assertions About Concepts bookmark33](#)
- ◆ the inlineRef property: helps to reference concepts which are inline, within free text properties of type label, see [References to Inline Concepts](#) (page 30)
- ◆ the derivedFrom property: helps to document from which concept a concept, used as property value in this item, has been derived, see [Document Derivation of Concepts](#) (page 31)

9.6 Planning ItemContent

A **newsCoverageSet** wrapper element contains a set of newsCoverage components (see below). Their order of appearance in conceptSet is not relevant. The major reason for having multiple newsCoverage components in this set is that each newsCoverage may be bound to e.g. a specific itemClass. Thus for including the coverage of an event by two text stories, 10 photos and one graphic one would have to use three newsCoverage components.

The **newsCoverage {Planning}** component holds the mandatory planning property and the optional delivery property.

The **planning** property - at least one has to be present in the wrapper - provides a rich set of properties which should tell the receiver what kind of coverage he can expect from the provider:

The **g2contentType** and the **itemsClass** properties tell what type of G2 deliverables to expect, and the **itemCount** adds how many of them to expect. The properties **scheduled** and **service** add when and by which service, or feed, the coverage will be delivered. A group of Descriptive Metadata gives a hint for the metadata which will be used with the delivered items allowing the receiver to build a filter or to forward this planning information to the proper desk of his editorial. The **assignedTo** property holds the person, organisation or company who has to deliver the content, this property can be of internal use of the news provider only or may be used to let the receivers know that e.g. a well-known journalist will write a review of a cultural event. If anything cannot be expressed by these machine readable properties the **edNote** may be used to add more information by natural language.

The **delivery** property can add what of the planned coverage has been delivered. It is a wrapper for a set of **deliveredItemRef** properties, each of them pointing to a G2 property which has been delivered.

Be aware that the **itemMeta** wrapper of all G2 items includes a **deliverableOf** (page 147) property. This property can be considered to be a link back to this Planning Item and a specific News Coverage component of it. By these means the receiver can check by using the **deliveredItemRef** properties if an item indicated as "being delivered" has already arrived. On the other hand if a G2 item being a deliverable of a planned coverage arrives before an updated version of the planningItem has arrived already indicates that it pertains to a specific editorial planning. A news provider should take care to keep these two links in synch.

9.7 Processing Considerations

It can be expected that Planning Items will have a high frequency of updates. The first version may be sent when the first outline of covering an event or a topic has been completed by the editorial of the news provider. Updates could and should be sent when types of planned G2 items extend (e.g. first text only is planned, later extended to text plus photos), when the number of planned items change or when typical metadata values for the items have been assigned. In the course of creating and delivering the items the Planning Item may be updated each time such an item (or a group of items, like a group of graphics) has been released.

10 Dealing with Controlled Values

10.1 {scheme, code} Pair, Scheme URI and Concept URI

Many properties usually have their value taken from a well defined scheme, i.e. a controlled vocabulary (i.e. a classification system, authority list, taxonomy, thesaurus etc ...).

These values are represented by a formal combination - a {scheme, code} pair - primarily intended to be consumed by processing software. A scheme is logically a closed set of related concepts, and a {scheme, code} pair unambiguously identifies a single concept.

A scheme is in practice a list of codes managed by a specific authority (which we shall refer to as the Scheme Authority), which may be the IPTC or any other well known standardisation body, or may be an



individual news provider or knowledge management company. A {scheme, code} pair therefore fully identifies a term from a controlled vocabulary (i.e. a scheme). A code **MUST** be persistent over time in order to avoid ambiguities when processing archives documents.

A scheme is fully and unambiguously identified by a scheme URI. The concept represented by a code is fully and unambiguously identified by a concept URI. The concept URI is obtained by appending the code to the scheme URI. **Qualified Code (QCode)** shows how a more compact form of a concept identifier is used in the news workflow.

As an example, an IPTC scheme for news categories might be identified by the URI "<http://cv.iptc.org/newscodes/mediatopic/>". If the code "15000000" represents the concept of "Sport", then the concept URI for "Sport" would be "<http://cv.iptc.org/newscodes/mediatopic/15000000>".

It is not mandatory that the Scheme Authority maintains the complete list of codes making up a given scheme in any particular form, e.g. as an XML file. It is sufficient that an unambiguous identifier is defined for each scheme a provider uses, and that this identifier is known by a Catalog (see **Catalog of Controlled Vocabularies**) to the customers of the news feed this provider offers.

Common needs are:

- ◆ To access human readable information about a scheme.
- ◆ To retrieve all terms of a scheme (e.g. to display a list of choice).
- ◆ To access human readable information about a qualified code.
- ◆ To check that a qualified code belongs to a scheme.
- ◆ To retrieve the definition of the concept identified by a qualified code in a given scheme.

Therefore the scheme URI **SHOULD** resolve to a web resource (or resources) containing information about the scheme in both human-readable and machine-readable forms. Meeting this requirement is mandatory for schemes which are to be compliant with the Semantic Web.

The concept URI **SHOULD** also resolve to a web resource (or resources) containing information about the concept in both human-readable and machine-readable forms. Meeting this requirement is mandatory for concept URIs which are to be compliant with the Semantic Web.

If content negotiation is implemented using HTTP, then the HTTP Accept header should be used to request information in the required format and the HTTP Accept-Language header should be used to request information in the required human language.

When designing a scheme URI, the following points should be taken into consideration:

- ◆ Each scheme URI must end with a suitable terminating character, e.g. "/" or "#". Each of these has various advantages and disadvantages, which are discussed extensively in documents available on the Web.
- ◆ One point worth mentioning here is that not all strings which can be used to construct a legal URI are automatically legal in the context of HTML. For example, "<http://cv.iptc.org/newscodes/theme.html#15000000>" is not a legal HTML URI, as an HTML fragment name cannot start with a digit.

10.2 Qualified Code (QCode)

In order to manipulate controlled values in an efficient manner, a compact representation of a concept identifier is needed, a syntax which allows the use of a {scheme, code} pair as the value of an XML attribute.

For this purpose a short string called scheme alias (aka prefix) is defined by a provider as a substitute for a scheme URI in the scope of a single Item, and a compact notation of a scheme-code pair is defined, called Qualified Code or QCode.

The datatype for a compact notation of a scheme-code pair is called qualified code or more simply QCode. QCodes are the mandatory way to express controlled values in properties like **itemClass** (page 201) or **pubStatus**.

QCodes are notated by the following syntax: a scheme alias acting as a first part, followed by a colon (:) character, followed by a code from the scheme. They are case sensitive.



The value space of the **QCodeType** datatype is a set of {scheme, code} pairs which identify concepts.

Note that:

- ◆ This is similar to the value space of the QName datatype, i.e. a set of {namespace, local part} pairs which identify element or attribute names.

Note: QNames cannot be used for this purpose, because the local part of QNames cannot be a numeric, but the News industry and the Financial industry are full of taxonomies making use of numeric codes. They aren't alone in this aspect (consider ISBN and ISSN).

- ◆ QCodes allow any sequence of legal URI characters in the code part, including, for example, digits only, dashes, slashes, etc.
- ◆ QCodes MUST have a non-empty scheme alias.

QCodes can be viewed to a certain extent as short, lexical representations of URIs. Be careful: the mapping from a qualified code to a URI is not bijective: a URI cannot be mapped back to a qualified code, because the separator of the tuple is not explicitly defined in the URI.

In order to resolve a qualified code, a processor MUST loop through the **scheme** elements defined in the scope of the Item. If the QCode scheme alias is found as value of the **alias** attribute of a scheme element, the scheme URI is the associated **uri** attribute and the controlled value is the resulting {scheme URI, code} pair. If no corresponding scheme alias is found, the processor SHOULD raise an error and consider that the property has no value.

10.21 Lexical Space Specification and Processing Model for Scheme URIs, Scheme Aliases, Codes, and QCodes

10.2.1.1 Lexical Space

- ◆ Lexical space for scheme URIs: conforms with the Unreserved Characters of RFC 3986, section 2.3. Reserved Characters as per RFC 3986, section 2.2 must be considered depending on the selected URI scheme.
- ◆ Lexical space for Aliases: all characters except colon (#u003A) and white space (#u0020 | #u0009 | #u000D | #u000A).
- ◆ Lexical space for Codes:
 - All Unreserved Characters of RFC 3986, section 2.3.
 - Reserved Characters as per RFC 3986, section 2.2 must be considered depending on the selected URI scheme. See also section 10.2.1.2 (Creating Codes) below.
 - As an alternative to percent-encoding whitespace characters (#u0020 | #u0009 | #u000D | #u000A) as defined by RFC 3986, these characters may be replaced by a sequence of one or more unreserved characters - like e.g. underscore or hyphen - 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 in that scheme

10.2.1.2 Processing Model

10.2.1.2.1 Creating Scheme URIs

Define a URI complying with the rules defined in 10.1 and 10.2.1.

Note: every scheme URI must comply with the RFC defining URIs (3986) or IRIs (3987).

10.2.1.2.2 Creating Codes

As defined in section 10.1 a concept URI is created by appending the code of a concept to the scheme URI of the vocabulary.

Therefore, appending a Code to a valid Scheme URI must make a valid URI, in particular, the Code must only contain characters that are legal URI characters (RFC 3986). As defined in RFC 3986, the Scheme Authority MAY percent encode reserved characters as well as the percent ("%") character depending on the role of each character as defined by that specific publisher for this Concept URI. Any percent encoding which is applied to characters in the code of a Concept URI MUST be used also by the corresponding QCode.



The Scheme Authority is fully responsible that concept URIs defined by him are properly resolved by any resolution system provided by him.

The examples below show how to deal with a string which should be used as code and includes a reserved character:

Example **without** encoding a slash in the code:

String to be used as Code: ebc13/14

Code: ebc13/14

Scheme URI: <http://cv.example.org/schemeA/>

Scheme Alias: schA

QCode: schA:ebc13/14

Concept URI: <http://cv.example.org/schemeA/ebc13/14>

Note: The Scheme Authority has to take care that this URL is resolved properly by a resolution system provided by him.

Example **with** encoding a slash in the code: String

to be used as Code: ebc13/14

Code: ebc13%2F14 (with applied percent encoding)

Scheme URI: <http://cv.example.org/schemeB/>

Scheme Alias: schB

QCode: schB:ebc13%2F14

Concept URI: <http://cv.example.org/schemeB/ebc13%2F14>

10.2.1.2.3 QCodes

10.2.1.2.3.1 Creating QCodes

Concatenate the Scheme Alias, a colon and the Code to form a QCode.

10.2.1.2.3.2 Inserting a QCode as the value of an attribute in a G2 XML document

1. Take a QCode as created in 10.2.1.2.3.1 and apply any required XML encoding to this string (Note: this is typically done by the XML processor software).
2. Insert the resulting string into an attribute as the QCode value.

10.2.1.2.3.3 Receiving/Parsing QCodes from an XML Document (any G2 Item)

1. Retrieve the QCode value from the XML document
2. Apply any required XML decoding (Note: this is typically done by the XML processor software).
3. To split a QCode into a Scheme Alias and a Code, identify the first colon, searching from left to right. The string to the left of the colon is the Scheme Alias; the string to the right is the Code. If no colon is found, the QCode is invalid.
4. Check whether the alias is defined in the catalog. If it is not, the QCode is invalid.

10.2.1.2.4 Concept URIs

G2 processors should be able to process Internationalized Resource Identifiers (IRIs) as per RFC 3987.

10.2.1.2.4.1 Creating a Concept URI/IRI:

Concatenate the Scheme URI and the Code to obtain the Concept URI.

10.2.1.2.4.2 Comparing Concept URIs/IRIs:

If provided Concept URIs are IRIs per RFC 3987 then they must be compared for equivalence as defined per RFC 3987, section 5.

If provided Concept URIs are URIs per RFC 3986 then they must be compared for equivalence as defined per RFC 3986, section 6.

10.3 Processing Catalogs

10.3.1 Structure of a Catalog

A **Catalog** MUST have one or more scheme elements. A catalog MAY have one or more titles in different languages. It MAY also have a pointer to some additional information available on the Web, and especially its



evolution by identifiers of a web location from where it can be retrieved, an identifier of the catalog and the version, and an identifier of the authority which manages this catalog. Such information will help people follow the evolution of a shared catalog like the IPTC G2 catalog, and include in their Items a reference to the latest version if they wish. A catalog may be managed by a provider by using a Catalog Item (available only at PCL) - see [Managing Catalogs - catalogItem](#) .

It MUST have one or more [scheme](#) elements. A scheme element MUST have a scheme *alias* attribute and a corresponding scheme *uri* attribute. It MAY have a name, a definition and a note element to provide human readable information about the scheme. And the authority governing the scheme MAY be indicated by the authority attribute. A [sameAsScheme](#) element MAY be used by applying a URI which identifies another scheme with concepts that use the same codes and are semantically equivalent to the concept of this scheme.

Each instance of an Item defines its own set of scheme definitions, and there is no interaction between scheme definitions in different Items. Scheme alias declarations are local to the Item in which they appear and cannot be overridden in a given Item.

10.3.2 Processing Remote Catalogs

By activating the hyperlink of a remoteCatalog (see [catalogRef](#)), a plain catalog structure is returned, and MUST be processed as if were locally defined.

10.3.3 Caching a Catalog

It is recommended for a processor to cache a remote catalog indefinitely, so that provider's servers are not overcharged with file requests.

When a processor opens an Item, it MUST check the URL(s) of the catalog(s) found in its header. If a catalog has not been previously cached, the processor MUST fetch it, check it, and SHOULD store its content in a cache.

Different remote catalogs MAY define different mappings for a given scheme alias. An entry in a remote catalog cache is therefore a triple {remote catalog URL, scheme alias, scheme URI}.

10.3.4 Checking a Catalog

It is OK for one scheme URI to have two aliases. It is an error if one alias is mapped to two different URIs in the scope of a single Item (an issue called alias collision). Note that this error may arise within a catalog, as well as across a set of catalogs (local or remote) declared in a given Item.

Before processing an Item, a processor MUST check its catalogs. If an alias collision is found, the processor MUST reject the Item as it can lead to misinterpretation of the information.

Note: If an aggregator finds an alias collision (i.e. the same alias associated with two URIs) while creating a packageItem which aggregates content from various providers, the aggregator MUST change one or both of the aliases before publishing the packageItem. This can be done by creating and publishing one or more non-clashing external catalogs (which replace the original external catalogs) and/or by replacing one or more external catalogs with non-clashing in-line scheme declarations.

10.4 Processing Schemes

10.4.1 Evolution of Scheme URIs

Schemes evolve: terms are added, names are changed, terms are retired. An authority will release a new version after each update. A provider may not want to adopt the latest version of a scheme. The scheme URI MUST be stable as long as the evolution does not break backward compatibility rules.

10.4.2 Retrieving All Terms of a Scheme

Here we are interested in schemes defined as an explicit list of terms. Schemes defined via an algorithm are out of scope of this section. A scheme definition is defined as the finite set of terms composing a scheme. A scheme definition MAY be a subset of an original scheme, e.g. maintained by an external authority.

Note: An authority is not necessarily able to make scheme definitions available for operational use, and a provider may use only a subset of the scheme defined by an authority.

A provider SHOULD make a scheme definition available for its users for operational use as the content of a



knowledgeItem, where each term is represented as a concept component, i.e. a concept identifier, a list of names in one or more languages, plus additional properties of the concept (all but the identifier being optional).

An authority MAY provide different variants of a scheme definition, e.g. a list of codes, a list of codes plus a name in a specific language, a list of codes plus names in all available languages.

For each variant of a scheme definition, the URL of the corresponding knowledgeItem SHOULD be available using e.g. content negotiation.

Selection from among the renditions MAY be performed automatically (if the processor is capable of doing so) or manually by the user selecting from a hypertext menu.

10.5 Qualified and Typed Properties

Qualified properties – of datatype **QualPropType** – only support controlled values in the short format of QCodes or full URIs.

Rule for using a QCode (qcode attribute) and a full URI (uri attribute) in property:

- An element SHOULD NOT use both a @qcode and a @uri. This rule applies to all properties except conceptId.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence.

A large subset of these properties supports concepts of different types as a value. Therefore typed qualified properties – of datatype **TypedQualPropType** – additionally provide a concept type relative to the value of the property.

For example, the type of the concept assigned as subject of a news story may be a theme (e.g. sport or football), a person, an organisation, a geographical area, a point of interest, an event, a business sector, a currency etc. The concept type of a **creator**, **contributor** and **infoSource** (page 200) of an Item may be a person or an organisation

Qualified properties MAY be complemented by one or more names associated with the underlying concept. Names can be expressed in different languages or variants.

10.6 Flexible Properties

It is not always possible or sensible to use a concept identifier (either as QCode or full URI) as metadata value. As an example, few news organisations maintain a formal listing of their editors, and therefore using a controlled value for the creator property is not always possible.

In order to fulfil this need, a large number of properties allow that literal identifiers or no identifiers at all to be applied instead of controlled identifiers. Additionally, a free-text value in the literal attribute is an identifier of a concept and NOT a human readable description. Therefore flexible properties - of datatype Flexible Property Type or a derived datatype - support both controlled (qcode or uri) and uncontrolled (literal) identifiers - or no identifier at all.

QCodes or URIs on one side (find more about their use in **Qualified and Typed Properties**) and literals on the other are mutually exclusive for any given property; if one of them exists the other one MUST NOT exist. (The term qcode/uri below indicates that the qcode or the uri or even both attributes can be used to express a controlled value.)

The rules for using the qcode/uri or the literal attribute or no concept-identifying attribute at all with a property are:

- If a bag is used with a property then qcode/uri and literal attributes MUST NOT be used with the property.
- If a bag is not used with a property then the property MAY have a qcode/uri attribute OR a literal attribute or neither.
- If a literal is used with an assert property then all instances of that literal in that item MUST identify the same concept.
- If a literal is not used with an assert property then it is NOT required that all instances of that literal in that item identify the same concept.



Literals MAY be used in the following cases:

- 1) As an identifier for linking with an assert element inside a G2 item: The value could be a random one. If a literal value is used with an assert property then all instances of that literal value in that item must identify the same concept.
- 2) When a code from a vocabulary which is known to the provider and the recipient is used without a reference to the vocabulary: The details of the vocabulary are communicated outside of the G2-Standards specifications. Such a contract could express that a specific vocabulary of literals is used with a specific property.
- 3) When importing metadata: The values of literals may contain codes which have not yet been checked to be from an identified vocabulary.

The value of a flexible property identifies a given concept with a specific type. It is useful to express e.g. that the provider of a news item is a person or an organisation. The *type* of the concept MAY be indicated as an attribute of the flexible property.

One or more additional name properties MAY be provided in different languages and variants for display. If the value of the property is a literal and no additional name is given, the recipient MAY use the literal value for direct display. But as the primary use of a literal is being an identifier it may not tell much about the meaning of the metadata.

Flexible properties MAY also be complemented by other information about the concept, like properties from Concept Relationships Group (see [Table 229](#)) and Concept Definition Group (see [Table 228](#)).

Flexible properties which value specifically identifies a person, an organisation or any other entity for which detailed properties are defined in this specification, MAY contain detailed information about this entity, e.g. a date of birth for a person or a location for an organisation.

Such information constitutes “hints” about the concept, which may be useful for display or indexing, but which should not be used to convey knowledge stored as-is in a knowledge repository. A specific mechanism, based on conceptItems and knowledgeItems, is set-up in the News Architecture for managing knowledge.

10.7 Composite Concepts

Several flexible properties support composite concepts. Composite concepts, a.k.a. pre-coordinated terms, are “glued” together to represent a concept made of atomic parts.

Therefore flexible concept properties – of datatype [Flexible 1 Concept Property Type](#) – have a bag child element which is used to express a new concept, composed from multiple existing concepts. The description of each existing concept is placed in a bit child element of the [bag](#) wrapper.

Examples of possible composite concepts are:

- ◆ John Doe Smiling {John Doe + Smiling }
- ◆ Women's 100m Swimming Final {Women + Swimming + 100m + Final}
- ◆ Positive pre-announcement by Citigroup {Citigroup + Pre-announcement + Positive}
- ◆ Microsoft's share price has moved up {Microsoft + Share price + Up}
- ◆ The Clintons {Bill Clinton + Hillary Clinton}

10.8 Editing Attributes

In a professional and collaborative news workflow, it makes sense to identify all elements defined by the model in order to later act on them individually. Also, metadata is not always entered by one person at one time, but may be entered by different people, organisations or systems at different time. Therefore it may be needed to keep track of who is assigned the editing of specific properties, and when and by whom a property has been given a value.

For this purpose, all metadata properties share the Common Power Attributes Group (see [Table 279](#) on page 361), which includes an optional local identifier (id) and the optional indication of the creator and the date (and, optionally, the time) when the property was last modified. (Beyond that the group includes more



attributes for other purposes.)



11 Managing Catalogs - catalogItem

An XML Schema file corresponding to the specifications for this item is available (see [The Full Set of Specification Documents](#) on page 3).

11.1 Description

Catalogs have a quite vital role for all the different NewsML-G2 item types as they provide a key resource for resolving QCodes to URIs identifying concepts: the mapping between scheme URIs and scheme aliases. This is explained in depth in [Processing Catalogs](#) (page 55).

In this context some providers may wish to use the same basic means for managing a Catalog as are available for news content, concepts, editorial planning etc. This purpose is covered by the Catalog Item which has been introduced in NewsML-G2 2.15.

A Catalog Item enables to user to cover this kind of catalog management:

- ◆ A specific set of [Scheme Declaration](#) elements is considered to form a catalog.
- ◆ This catalog is made available by a single catalog element which may appear in a stand-alone file as web resource or which may be included into NewsML-G2 items.
- ◆ This catalog element can be made the content of a Catalog Item.
- ◆ The scheme elements of this catalog may be changed (modified or a new one added), but by the general rules of NewsML-G2 this requires to create a new version of the Catalog Item.

11.2 Structure of a Catalog Item

The model of a catalog item is very similar to the other G2 items: It shares the indicators of compliance with a standard and a conformance level, Identification and Versioning, Signature, Rights Information, Item Metadata and Item links. Please review [Representing News - newsItem](#) for more information.

11.3 Item Class

The IPTC provides a mandatory standardised scheme applicable to the itemClass property of a catalogItem, identified by the URI <http://cv.iptc.org/newscodes/catinature/>.

11.4 Catalog Related Metadata

The set of metadata related to the catalog-content is listed below. All properties are optional. The order of the properties in this set is flexible: the non-repeatable properties MUST come first and then the repeatable properties may be inserted in any order.

Table 11. Content Metadata Elements

Element Title	Element Name	Card	Described on Page
Date Content Created	contentCreated	(0..1)	129
Date Content Modified	contentModified	(0..1)	130
Creator	creator	(0..unbounded)	125
Contributor	contributor	(0..unbounded)	120
Alternative Identifier	altId	(0..unbounded)	73

Each provider may add a set of metadata properties which have to be defined in a non-IPTC-G2 namespace. See also [XML Namespaces](#) and [Extension Points in XML](#).

11.5 Catalog Item Content

A Catalog Item includes a mandatory Catalog Container element which contains a single and mandatory [Catalog](#) element as the content of the item.



12 Dealing with Labels and Blocks

12.1 Introduction

Labels expose aspects of news as natural language strings. They are assertions expressed as natural language strings intended to be consumed by human beings. They are typically displayed alongside the content of an Item or in place of Items in a list, providing a means of selection among them.

Blocks are simply labels with an additional line break. They are primarily used for notes, comments or instructions created by a news provider for use by recipient editorial teams.

Labels and blocks MAY have a **role** attribute, which refines the semantics of the property.

Labels and blocks MAY have a **media** attribute. When present, the value MUST conform to the Cascading Style Sheets specification [CSS]. Several media types can be given as space separated values.

All labels and blocks support rich text, i.e. text interspersed with some specific markup, identical to XHTML1.1 markup: the anchor (**a**) for the inclusion of hyperlinks, the **span** as a generic mechanism for adding information to text, simple **ruby** markup used in Japanese publications and **inline** for semantic inline markup.

The inline property identifies a concept present in a label or block either by a qualified code or a literal value, plus an optional type. Additional information about this concept can be represented using an **assert** property value, plus a basic set of properties defining the concept.

12.2 Internationalization Attributes

In an international news workflow, fine grained control of language information in the hierarchy of nodes constituting an Item is needed.

For this purpose, all labels – and all ancestors of such an element – share an International Attributes Group (see [Table 278](#)), i.e. an optional language tag (xml:lang) and indication of the directionality of textual content (dir).



13 Exchanging Items - newsMessage

An XML Schema file corresponding to the specifications for this news message is available (see [The Full Set of Specification Documents](#) on page 3).

13.1 Description

A **newsMessage** facilitates the exchange of all kinds of items by any kind of digital transmission, especially in a broadcast or multicast network.

The content of a newsMessage is an **itemSet** component.

An itemSet contains a set of newsItems, packageItems, conceptItems and knowledgeItems. The model assigns no significance to the order of Items within the News Message.

The use of a News Message is totally optional in a news workflow. Alternatively, Items may be exchanged using SOAP, WebDAV, ICE, the Atom Publication Protocol (using Atom feeds, and items as payload of an Atom entry) or any other possible syndication protocol.

It may be useful for a recipient to store the information conveyed by a message, but this is not mandatory. Usually the messaging information will be maintained separately from the information relative to the contained items.

13.2 Message Information

All the information about the newsMessage as a wrapper of conveyed G2 items is collected under the **header** element which **MUST** be present.

A newsMessage **MUST** have a date of transmission – **sent** . The date of transmission **MAY** not be updated in case of retransmission of the message.

If any QCode is used within the header then a catalog and/or a catalogRef property **MUST** be included in the header. The scope of the scheme elements of the local and/or remote catalog(s) is limited to the header element and its descendants and explicitly does **NOT** extend to the children of itemSet.

A newsMessage **MAY** have a **sender** child element, which may be an organisation or a person. The structure of this string is not specified by the IPTC. Best practice is to identify a sender by its domain name.

It **MAY** have a transmission identifier –**transmitId** – and a priority of transmission. No two newsMessages sent by the same sender on the same date can have the same identifier. In case of retransmission it is not required to update this identifier. The structure of this string is not specified by the IPTC.

It **MAY** have a **priority** property to control the overall message transmission process. It **MAY** indicate the point of **origin** of the message, using a provider defined syntax.

It **MAY** have one or more **timestamp** (page 299)(s) associated with the message. The exact meaning of this timestamp may be refined by a **role** attribute.

It **MAY** have one or more **destination** properties using a provider defined syntax, and the indication of one or more channels – **channel {News Message}** – of transmission.

It **MAY** have one or more **signal** properties to instruct the news message processor that the content requires a specific handling.

Each particular provider is equally able to add to this set information of its own, by mutual agreement with the recipients of the Item.

13.3 About Using Schemes in a newsMessage

It is important to note that a newsMessage does not define any catalog that would be common to the Items it contains. There is no interaction between the scheme declarations present in different Items exchanged in a newsMessage.



14 Specification Reference

This section provides specifications to be combined with the NewsML-G2 XML Schema.

How to access the XML Schemas is defined in the [The Full Set of Specification Documents](#) section at the very start of this document.

14.1 Introduction to the Common Components

News exchange formats share many metadata properties as they are about the same data: something newsworthy to be exchanged. For that reason the family of IPTC G2-Standards share a large set of properties which are common to all family members and this common data model and set of specifications is called the IPTC News Architecture for G2-Standards (NAR).

This Specification Reference section provides a mix of specifications coming from the NAR and additionally from this G2-Standard.

The components specified in this Specification Reference can be split into these 3 groups:

1. Fine grained components, called a datatype. A datatype has no specific business meaning or semantics of its own and only takes on business meaning when used as the data type of a property. For NewsML-G2 the names of datatypes end with a "Type" suffix (e.g. QCodeType). Datatypes fall in two groups:
 - Simple data types are primitive data types, as found in software languages or XML schema definitions (e.g.. integer, string). Some restriction may be imposed, such as Int100Type where an integer has been restricted to a value range of 0 to 100.
 - Complex data types are simple data types extended to add further information in order to correctly represent the value. Such ancillary information takes the form of attributes. For example a LabelType supports mixed content and is extended with language and role attributes.
2. Medium grained components, called basic component or property. A property represents a single piece of business information and uses an existing data type or defines it own local datatype to provide its content model. It is capable of being used independently or as part of a group. Like a complex data type, a basic component can be qualified by ancillary data if required to complete its meaning. For example, a slugline element of data type string supports an additional separator attribute.
3. Coarse grained components, called aggregate component. It is a collection of properties that together is more than the sum of its constituent parts. The properties composing the whole can be properties or aggregate components. An aggregate component may be designed so it supports an extension point where news providers can extend its usage. For example, a descriptive component is defined as a group of properties like title and subject, and a person component is defined as a group of properties like name and date of birth.

14.2 General Specifications

14.2.1 XML Namespaces

Table 12. XML Namespace

Namespace URI	Recommended Alias	Usage Note
http://iptc.org/std/nar/2006-10-01/	nar	For all common components of the family of IPTC G2-Standards

14.2.2 MIME Types

Table 13. IANA Media Types (so called MIME Types)

IANA Media Type Identifier	Usage Note
application/vnd.iptc.g2.newsitem+xml	For all kinds of G2 News Items.



application/vnd.iptc.g2.conceptitem+xml	For all kinds of G2 Concept Items.
application/vnd.iptc.g2.packageitem+xml	For all kinds of G2 Package Items.
application/vnd.iptc.g2.knowledgeitem+xml	For all kinds of G2 Knowledge Items.
application/vnd.iptc.g2.planningitem+xml	For all kinds of G2 Planning Items.
application/vnd.iptc.g2.catalogitem+xml	For all kinds of G2 Catalog Items.
application/vnd.iptc.g2.newsmessage+xml	For the G2 News Message

All these Media Types are registered with IANA, see <http://www.iana.org/assignments/media-types/>

14.2.3 Extension Points in XML

For attributes: each element of a G2-Standard allows to add provider specific attributes from any other XML namespace than the News Architecture for G2 namespace.

For elements: Some elements which have child elements allow to add provider specific elements from any namespace other than the News Architecture for G2 namespace. A few elements allow adding any element from any XML namespace - including the News Architecture for G2 namespace - but this is a special case only, see below.

14.2.4 Hint and Extension Points in XML

To act as an Extension Point properties from any other XML namespace than the News Architecture for G2 namespace may be added.

To act as a Hint Point properties from the News Architecture for G2 namespace may be added.

The purpose of properties from the NAR namespace is to add a set of hints, i.e. properties which have to comply with the structure of the G2 item target resource but do not have to be extracted from it. These properties must be added this way:

- Immediate child properties of <itemMeta>, <contentMeta>, or <concept> - optionally with their descendants - may be used directly under the extension point
- All other properties require the full path excluding only the item's root element.

14.3 Implementation Design Rules

These design rules were applied while developing the G2-Standards. Some apply to all kinds of technical implementations, other only to one specific implementation.

- ◆ Each element supports a set of common attributes.
- ◆ Each element has an extension point at the attribute level (XML implementation only).
- ◆ Each element containing an international string supports i18 attributes.
- ◆ Each ancestor of an element containing an international string supports i18 attributes.
- ◆ Children of wrapper elements: mandatory children come first, they are in a specific order, optional (and in most case multiple) elements follow, they can be inserted in an arbitrary order (XML implementation only).
- ◆ Each wrapper element has an extension point as its last child element (XML implementation only).

14.4 Processing Model Terminology

For many components of the G2-Standards this specification provides also a processing model. Find below how these processing instructions should be read.

- ◆ A Processing Model provides rules for the proper processing of metadata properties and their values. Each rule may be divided into steps.
- ◆ Each rule gets an integer number assigned, steps for this rules are indicated as decimals to this number. Example: rule 12, step 3 = 12.3
- ◆ Many rules can be considered like a function in programming, hence as a sequence of processing steps in the scope of a block. These terms will be used for defining the rules and are based on this



basic layout:

- “quit” = the processing of this function stops at this step and quits the current context to the calling context.
- “quit and return ...” = see “quit”, plus: a value of “...” is returned to the calling context.
- “if ... :” = a condition is expressed and right to the colon the processing that results from meeting this condition.
- If the condition is NOT met the default processing is “proceed to the next step of this processing rule”. A specific processing for this case is preceded by the term “otherwise”.



14.5 Element specifications

All XML elements of NewsML-G2 and their structure are defined by the NewsML-G2 Schema - see how to get them in the [The Full Set of Specification Documents](#) section.

This section only adds for some elements User Notes, Implementation Notes or recommended Controlled Vocabularies which are not in the NewsML-G2 XML Schema version 2.23.

All elements not having such additional information are not listed here.

14.5.1 Accountable Person – accountable

User Note: This property answers to a legal issue. In some countries (e.g. Germany, Sweden) it is needed to designate a person accountable for any legal issue related to the published content. The full translation from the German term is: accountable person in terms of the press law - (For reference in German: Verantwortlich im Sinne des Presserechts -acronym = ViSdP), in Swedish it is called "Ansvarig utgivare". In practice today, a news provider may send out a message each day which indicates the "accountable person". This may work for traditional feed services, but fails with profiled services (content selections) which filter such messages. The solution is to include this information in the Items themselves.

14.5.2 Action in Hop History – action

Recommended IPTC NewsCodes CV for the target attribute: <http://cv.iptc.org/newscodes/hopactiontarget/>

14.5.3 Alternative Identifier – altId

User Note: If there is more than one alternative identifier, they SHOULD be qualified using the type qualifier to distinguish between different identification schemes.

14.5.4 Alternative Locator - altLoc

User Note: If there is more than one alternative locator, they SHOULD be qualified using the type attribute to distinguish between different identification schemes.

14.5.5 Alternative Representation - altRep

User Note: This property is particularly useful if the Item is available in different formats (for example NewsML 1, IIM or NITF) or with different levels of details (for instance with different granularity of metadata).

14.5.6 Assertion - assert

User Note: The assertion about the concept may be used to merge multiple occurrences of concept details in properties into a single place or to extend the details of an assertion beyond the limited details other properties can provide.

Rule for @qcode and @uri in an element:

- An element SHOULD NOT use both a @qcode and a @uri.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence

14.5.7 Bag Item – bit

Note/example:

Note on the significance attribute:

This attribute is assigned to a special use case of a bag with subject properties: the bag includes one bit representing an event and one or more other bits representing entities which are related to this event. Only in this case the significance attribute may be used to express the significance of this event to the concept of the bit carrying this attribute. If the bag includes more than one event, any significance attribute of bits in the bag SHALL be ignored.

Example 1:

A merger of two companies which is differently significant to the two parties of the merger: the significance of the merger for the small company is high while it is low to the global player company.

```
<bag>
  <bit type="cpnat:event" qcode=" abevents:Merger123AB"/>
  <bit type="cpnat:organisation" qcode="isin:TinyCompany"
significance="100"/>
  <bit type="cpnat:organisation" qcode="isin:GlobalPlayerCompany"
significance="10"/>
```



</bag>

14.5.8 Broader – broader

User Note: rank is suitable for use in a Knowledge Item representing a scheme. It is used when it is important that the Child Elements of a particular term are displayed in a user interface in a predefined order.

For example, the major currencies could be given a rank of “1”, while all other currencies could be given a rank of “2”. Terms of the same rank are ordered alphabetically by name if this is available. If the name is not available, the terms are ordered by code value.

Terms without a rank are treated as if they all have the same rank, which is higher than the rank of all other terms.

The same concept may have different ranks in different concept trees. A lower rank results in a placement earlier in a display.

14.5.9 By - by

User Note: The by label provides a natural-language statement of the author/creator information (commonly called the byline); it may include a byline title, i.e. the author's job title.

Examples of bylines are RUPAK DE CHOWDHURI (a person), isotype.com (a provider) or STR (a stringer). It is up to the provider to decide if the label starts with a word like “By”.

14.5.10 Channel for News Message – channel

User Note: A channel identifier is used to provide recipients with information for selecting, routing, or handling otherwise the content of the message. The channels represent streams in a multiplex: a message may be sent on different channels – e.g. one for text, one for pictures – and each reception point will be able to filter on channel values. The structure of this string is not specified by the IPTC.

Rule for @qcode and @uri in an element:

- An element SHOULD NOT use both a @qcode and a @uri.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence.

Implementation Note: If both are present the @literal and the property string value SHOULD be identical. If both are present but not identical @literal takes precedence.

14.5.11 Circle – circle

User Note: The position element defines the centre of the circle.

Example:

```
<geoAreaDetails>
  <circle radius="1.335" radunit="dimensionunit:km">
    <position ...>
  </circle>
</geoAreaDetail>
```

14.5.12 Concept Definition – definition

User Note: A natural-language definition of the semantics of the concept. This definition is normative only for the scope of the use of this concept.

14.5.13 Concept Name – name

Recommended IPTC NewsCodes CV for the *part* attribute: <http://cv.iptc.org/newscodes/namepart/>

14.5.14 Confirmation – confirmation

Recommended IPTC NewsCodes: <http://cv.iptc.org/newscodes/eventdateconfirm/>

14.5.15 Contact Information – contactInfo

User Note: The *role* attribute addresses the role of the full set of contact information with regards to the entity defined by the concept. Examples: “privateOffice” vs “companyOffice” or “GlobalHeadquarters” vs “localHeadquarterUK”.

Recommended IPTC NewsCodes for the “role” of an event's contact information: <http://cv.iptc.org/newscodes/eventcontactinfoforole/>

14.5.16 Provider – provider

User Note: This property corresponds to the publisher of the Item.



14.5.17 Date Content Created - contentCreated

User Note: In the case of a photo or live footage for audio and video, this date (and time) is always the same as the date (and time) of the event covered by the content. In the case of text and any audio and video report about an event, this date (and time) can be different from the date (and time) of the event covered by the content. This date (and time) may also be different from the date (and time) of the creation of an Item holding the content.

14.5.18 Date Content Modified – contentModified

User Note: The value of this property should be updated each time the content is modified in any manner, but should not be updated if only metadata are changed.

14.5.19 Date of Transmission – sent

User Note: May not be updated in case of retransmission of the message.

14.5.20 Dateline – dateline

User Note: The dateline provides a natural-language statement of the date and/or place of the news content creation, to be displayed in situations where an abstract of the content is shown (case of search results) or the content is remote.

Traditionally a dateline indicates when and where news content is created, not necessarily the time and place relative to the news event.

As an example a dateline BAGHDAD, March 26, 2007 (AFP) could head a story about blast in Mosul, because the story was actually written in Baghdad. Also, by tradition a dateline will follow the stylebook of the information provider and possibly leave out certain time and location information that could be useful for specifying searches of a database. Editorial policy dictates the dateline; it is not automatically derivable from other markup (location, date, etc.). The dateline should not end with a separating character (of the kind that separates the dateline from the first sentence in a traditional wire story).

14.5.21 Description – description

Recommended IPTC NewsCodes for the *role* attribute: <http://cv.iptc.org/newscodes/descriptionrole/>

14.5.22 Destination – destination

User Note: In a broadcast delivery system, the destination is a group of reception points (using a provider-specific syntax, often geographically oriented). This is a way to address customers.

Examples are “England”, “USA”, “Austria/Vienna”, “France/Paris/LeParisien”.

The structure of this string is not specified by the IPTC.

Rule for @qcode and @uri in an element:

- An element SHOULD NOT use both a @qcode and a @uri.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence.

Implementation Note: If both are present the @literal and the property string value SHOULD be identical. If both are present but not identical @literal takes precedence.

14.5.23 Editorial Service – service

User Note: The values of this property are defined by each provider, and are often associated with the notion of a desk or a feed. Some examples are a “French wire service”, an “international picture service” or a “mobile news service”.

14.5.24 Event – event

Implementation Note: This event structure is used within an events wrapper to be plugged into an inlineXML property of a News Item.

14.5.25 Events – events

Implementation Note: This events wrapper is made to be plugged into an inlineXML property of a News Item.

14.5.26 G2 Content Type – g2Contenttype

Any of the G2-Standards specific IANA Media (MIME) Types like application/vnd.iptc.g2.*item+xml.

See: <http://www.iana.org/assignments/media-types/application/>

14.5.27 Generator Tool – generator

User Note: Where a role IS NOT specified, the Generator Tool applies to the most recent item generation stage.



Where a role IS specified, the Generator Tool applies to the item generation stage identified by the role.

14.5.28 Geographic Position – position

User Note: These properties follow the syntax used by the major geocoders on the Web. Latitudes north of the equator shall be designated by use of the plus sign (+), latitudes south of the equator shall be designated by use of the minus sign (-). The equator shall be designated by use of the plus sign (+).

Longitudes east of Greenwich shall be designated by use of the plus sign (+), longitudes west of Greenwich shall be designated by use of the minus sign (-). The Prime Meridian shall be designated by use of the plus sign (+). The 180th meridian shall be designated by use of the minus sign (-).

The altitude is given in meters. A positive integer means a position above the zero elevation, a negative value below the zero elevation. In the absence of the gpsdatum attribute, WGS84 is the default system

14.5.29 Group – group

User Note:

- Group Mode: By default the group is “complementary and unordered”.
- Complementary and Unordered: To be used for any kind of supporting content that does not require a sequence to be specified.
- Complementary and Ordered: The group starts with the first child of the group. To be used for any kind of content which must be displayed or consumed in a particular sequence, expressed by the order of the child elements of the group. The semantics of the role attribute value determine the required processing.
- Alternatives: To be used if a group contains equivalent pieces of content (e.g. translations of the same news story into different languages). The recipient may pick one or more of these.
- Group References and Item References: Can be included in any order, and this order may be relevant or not, depending the value of the mode attribute. Each link aggregates an external resource (Item or Web resource) to the package. Optionally, it indicates the relationship between the group and the target resource plus some additional hints about the resource itself.

14.5.30 Has Financial Instrument – hasInstrument

User Note: The symbolsrc and symbol attributes are a pair of values which define the authority which issued a symbol and the issued symbol.

The market can be defined in two ways: either by the market attribute which requires an identifier from a controlled vocabulary; or by a pair of marketlabelsrc and marketlabel values which define the authority which issued the marketlabel and the issued marketlabel

14.5.31 Hierarchy Info – hierarchyInfo

User Note: Represents the position of a concept in a hierarchical taxonomy tree by a sequence of QCode tokens representing the ancestor concepts and this concept.

Example: From the Media Topic NewsCodes (alias="mtp") using assumed codes:

The concept "adoption" has QCode mtp:2788

Its parent is the concept "family" with the QCode mtp:2780

The parent of "family" is the top level concept "society" with the Qcode mtp:1400

The resulting Hierarchy Info value is

```
<hierarchyInfo>mtp:1400 mtp:2780 mtp:2788</hierarchyInfo>
```

14.5.32 Hop – hop

User Note: The timestamp of the hop element reflects the time of forwarding the object while the timestamp of an action reflects the time of performing that individual action.

14.5.33 Inline Data – inlineData

Implementation Note: For the encoding attribute at the CCL only the QCode for “base64” may be used. If the attribute does not exist, this QCode must be assumed as default. In the absence of the encoding attribute, the content must be plain text, and the content type must be set accordingly.

14.5.34 Instant Messaging Address – im



User Note: The tech attribute indicates the provider of the service (Twitter, WhatsApp, etc.).

14.5.35 Information Source – infoSource

User Note: If no @role is applied the information source provided some information used to create or enhance the content and played no other role. Omitting @role is equivalent to applying <http://cv.iptc.org/newscodes/infosourcerole/origininfo> as the only role value.

If a party did anything other than originate information a role attribute with one or more roles must be applied. The recommended vocabulary is the IPTC Information Source Roles NewsCodes at <http://cv.iptc.org/newscodes/infosourcerole/>

To indicate that a party has modified or enhanced the content use the contributor property.

If an entity plays more than one role, the infoSource element has to be included multiple times, with different values of role.

Recommended IPTC NewsCodes for the *role* attribute: <http://cv.iptc.org/newscodes/infosourcerole/>

14.5.36 Item Class – itemClass

Mandatory IPTC NewsCodes for News Items or Package Items:

<http://cv.iptc.org/newscodes/ninature/>

Mandatory IPTC NewsCodes for Concept Items, Knowledge Items or Package Items:

<http://cv.iptc.org/newscodes/cinature/>

User Note: This property gives a hint on the nature of the Item. IPTC values for News Items correspond to the media type of the original content component, i.e. “text”, “photo”, etc. Concept Items adopt the static value *concept*. The class of a Package Item reflects the nature of the items it contains, i.e. either one of the values above or the value “composite” which indicates that the package handles items of different natures. A recipient system may use this information to make a coarse selection of Items, based on their nature, without having to inspect the structure.

14.5.37 Item Set - itemSet

XML Schema Notes: To allow the validation of the structure beyond the root elements of the different items the extension point “any” for the nar XML namespace is the only child element. This allows schema based validation of the content of the items as the validation of the extension point is set to “lax”.

14.5.38 Keyword – keyword

User Note: This property may be used in parallel with other properties to describe content like subject or genre which use QCodes or literals to identify an assigned concept.

Providers should define if and how the values of keyword properties contained in their items complement, or overlap with, the values of other properties such as subject or genre.

Implementation Note(s) Be aware of the lexical space restrictions for an XML Schema Normalized String type - see XML Schema specifications.

14.5.39 Language – language

tag values must be valid BCP 47 language tags.

Recommended IPTC NewsCodes for the *role* attribute: <http://cv.iptc.org/newscodes/languagerole/>

14.5.40 Line (of geoArea) – line

Implementation Notes: Order of positions has to be considered, a minimum of two position elements is mandatory

14.5.41 Link – link

User Note:

They are different variants of links:

Links may allow for navigation from a newsItem to another related Item or a Web resource, and its title be displayed as supplemental information to the final user. Example: a newsItem representing a section of a transcript (a “take” in the news language) may be linked to the previous and next take; an article about a person may be linked to the biography of this person.



Links may express a parent-child relationship. Example: a newsItem representing an article may be linked to the article it is a translation of; a wrap-up may be linked to the previous stories used as source material for the article; a cropped picture may be linked to its source picture.

Links may express dependency on external Items which are required in order to fully present the composite content of the Item. If some target Items are not retrievable, then the recipient processor should fail gracefully. The most obvious example is a newsItem representing an illustrated article. The textual content of the newsItem (usually formatted as NITF or XHTML) includes a reference to a photo which is represented by another newsItem. As the NAR recipient processor is content agnostic, it cannot infer this dependency from processing the content. A dependency link from the article to the picture indicates that the recipient processor must retrieve the target newsItem before the article can be fully displayed.

Pointing at the latest version of an Item while exposing content metadata may lead to unwanted display or selection criteria if these metadata were subsequently modified; therefore only the stable content properties should be exposed in a link.

14.5.42 Located – located

User Note: This information applies especially to news, and may also be expressed as free text in the “dateline” of a story, along with a date of content creation and the name of the content provider. The rules for determining the location are provider-dependent. The location is typically determined differently for different types of content:

- Text: The practices of news providers either identify the location the content relates to or the location the content was created by a reporter or a writer. If a correspondent is resident in town A but writes about an event in town B the name of town A or B can be used. But the provider's policy should be available as written document.
- Photo: The location of origin of content is the place shown in the photo image.
- Graphics: The location of origin of content should be the editorial office from where this graphics are distributed.
- Audio and video: In the case of raw footage the location of origin of the content should be the place of event, if people can be heard/are shown from different places the news provider can decide by its own policy, but this policy should be available as written document.

14.5.43 News Message Header – header

Implementation Notes: If any QCode is used within the News Message header then a catalog and/or a catalogRef element MUST be included in the header. The scope of the scheme elements of the local and/or remote catalog(s) is limited to the header element and its descendants.

14.5.44 News Content Characteristics – newsContentCharacteristics

Mandatory IPTC NewsCodes: <http://cv.iptc.org/newscodes/dimensionunit/>

14.5.45 News Coverage (of Concept Item) – newsCoverage

Implementation Note: Be aware that in EventsML-G2 version 1.6 this element was classified as LEGACY. From that version on a standalone Planning Item is available to hold an even extended set of information about planned coverage. Its major advantage is that coverage can be planned without having to update - and version - concept items for event concepts.

14.5.46 NewsCoverage (of Planning Item) – newsCoverage

User Note: A new newsCoverage property must be created for each set of planning details which contains different values. Different would be typically the g2contentType and/or the item-Class; or one or more of the descriptive metadata properties for the planned items.

14.5.47 News Coverage Status – newsCoverageStatus

User Note: Indicating a decision of coverage:

If a specific coverage was agreed by the news provider the newsCoverageStatus has to be set to code “int” (coverage intended) and at least one newsCoverage element with coverage details MUST be added to the eventDetails.

Highly recommended IPTC NewsCodes: <http://cv.iptc.org/newscodes/newscoveragestatus/>

14.5.48 Origin – origin



User Note: This string's structure is not specified by the IPTC.

Rule for @qcode and @uri in an element:

- An element SHOULD NOT use both a @qcode and a @uri.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence.

Implementation Note: If both are present the @literal and the property string value SHOULD be identical. If both are present but not identical @literal takes precedence

14.5.49 **Organiser – organiser**

Recommended IPTC NewsCodes for the *role* attribute: <http://cv.iptc.org/newscodes/eventorganiserrole/>

14.5.50 **Participant – participant**

Recommended IPTC NewsCodes for *role* attribute:

<http://cv.iptc.org/newscodes/eventparticipantrole/>

14.5.51 **Phone Number – phone**

User Note: The *tech* attribute indicates a land-line, cellular etc., service.

14.5.52 **Polygon – polygon**

Implementation Note: Order of positions has to be considered, a minimum of three position elements is mandatory

14.5.53 **Postal Address**

User Note: A special value of the *role* attribute may indicate that this information is not used to make contacts but e.g. is the registered address of a company.

14.5.54 **Postal Address of a Point of Interest – address (in POI structure)**

User Note: This address may be different from an address required to contact the Point Of Interest or the organisation running or maintaining it, that address is provided under a *contactInfo* element.

14.5.55 **Profile – profile**

User Note: This property gives information about the precise structure of an Item, e.g. a simple package, article with one picture, and may be the name of the transformation stylesheet used for the generation of the Item.

14.5.56 **Publish Status – pubStatus**

Mandatory IPTC NewsCodes: <http://cv.iptc.org/newscodes/pubstatusg2/>

14.5.57 **Rating – rating**

User Note: On @raters:

- a) If @raters is not present the number of raters defaults to 1
- b) @raters does not require that the count indicates distinct persons.

Implementation Note: on @valcalctype:

A CV for the calculation type should include: mean, median, sum, unknown

14.5.58 **Recurrence Group**

This group of properties defines the information required to specify a recurrence set. The recurrence set is the complete set of recurrence instances for a calendar component. The model follows the iCalendar specification [RFC2445].

At least one *rDate* or *rRule* element MUST be present. These elements MUST come first in the group. Then the *exDate* and *exRule* elements MAY be inserted in any order.

This group includes these elements:

- Recurrence Date – *rDate*
- Recurrence Rule – *rRule*
- Exclusion Date – *exDate*
- Exclusion Rule – *exRule*

14.5.59 **Registration - registration**

Recommended IPTC NewsCodes: <http://cv.iptc.org/newscodes/eventregrole/>

14.5.60 **Remote Content – remoteContent**

User Note: To identify the remote resource either the *residref* attribute or the *href* attribute MUST be



set, optionally both MAY be used in parallel. The residref attribute identifies a managed remote resource by its globally unique identifier (if the resource has such an identifier), while the href attribute identifies the location of the remote resource in e.g. a (remote) file system. If the remote resource is managed - like an item - and consequently the residref attribute is used, a version attribute MAY indicate the resource's version; in the absence of version information, the remote resource is the latest version available.

The Width Unit and Height Unit may take the following values, taken from an IPTC defined controlled vocabulary: lines, pixels, points (more units are defined by this CV, check the most recent version).

Mandatory IPTC NewsCodes: <http://cv.iptc.org/newscodes/dimensionunit/>

14.5.61 Role in the Content Stream – role

User Note: This property may indicate the role of the content part in a piece of streaming media.

Examples (video): “sting”, “slate”, etc.

14.5.62 Role in the Workflow – role

User Note: Among other possibilities this property may indicate the importance of the item in a fee by concepts like “flash”, “bulletin”, “alert”, “urgent”, “newsbreak”, and so on.

14.5.63 Ruby – ruby

Implementation Note: This implementation aligns with the Simple Ruby markup without and with parentheses of the W3C (see <http://www.w3.org/TR/ruby/#simple-ruby1>).

XML Schema Note: The alternative simple Ruby markup without and with parentheses is expressed by the use of either a single *rt* element or a single *rp-rt-rp* sequence of elements. Ruby parentheses (<rp>, empty elements) must be used as a pair: either both are present or none is present.

14.5.64 SameAs for a Scheme - sameAs

Implementation Note: This element SHOULD NOT be used in NewsML-G2 2.11 and higher, the element *sameAsScheme* should be used instead.

14.5.65 SameAs Scheme – sameAsScheme

Implementation Note: This element replaces the *sameAs* element as child of a scheme element.

14.5.66 Sender – sender

User Note: The structure of this string is not specified by the IPTC. Best practice is to identify a sender by its domain name.

Rule for @qcode and @uri in an element:

- An element SHOULD NOT use both a @qcode and a @uri.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence.

Implementation Note: If both are present the @literal and the property string value SHOULD be identical. If both are present but not identical @literal takes precedence.

14.5.67 Signal – signal

User Note: This property might indicate major rewriting of the content, important correction, urgent handling etc. The processor might be required to perform specific actions, depending on the contract between the provider and the recipient. Users should be alerted of the reception of an Item containing a signal by some UI mechanism (sound or display). An editorial note (edNote) may be used to convey additional natural language information related to the processing of the content.

14.5.68 Slugline – slugline

User Note: *separator* providers may choose to use more complex separation rules. In such a case the meaning of the separators must be conveyed by some other means

14.5.69 Time Delimiter – timeDelim

User Note:

The time unit may take the following values, taken from an IPTC defined controlled vocabulary:

- *timecode*: An SMPTE timecode containing a string encoded identification. Timestamp



format: hh:mm:ss:ff (ff for frames).

- *timeCodeDropFrame*: An SMPTE timecode containing a string encoded identification.

Timestamp format: hh:mm:ss:ff (ff for frames). The drop frame flag should be set.

- *editUnit*: The editUnit is the amount of time per video frame (1s / number of frames per second) or the amount of time per audio sample (1s / number of samples per second), for which the video frame rate or audio sample rate must be known. Timestamp format: long unsigned integer.

- *normalPlayTime*: Indicates the position relative to the beginning of the presentation.

Timestamp format: hh:mm:ss.mmm (mmm for milliseconds). See also: RFC 2326.

- *seconds*: Time given in full seconds. Timestamp format: long unsigned integer.

- *milliseconds*: Time given in full milliseconds. Timestamp format: long unsigned integer.

Implementation Note:

If a time unit IS NOT present, the value *editUnit* MUST be assumed.

Any timestamps MUST be formatted appropriately for the time unit (as detailed above).

All timestamps SHOULD be zero-padded from the left as applicable, e.g. a normalPlay-

Time value starting at 12 seconds would be represented as '00:00:12.000'.

Mandatory IPTC NewsCodes: <http://cv.iptc.org/newscodes/timeunit/>

14.5.70 Transmission Identifier – transmitId

User Note: This string's structure is not specified by the IPTC. No two News Messages sent by the same sender on the same date may have the same identifier. In case of retransmission it is not required to update this identifier.

14.5.71 Usage Terms – usageTerms

User Note: This property includes the type of usage to which the rights apply, the geographical area or areas to which specified usage rights pertain, the indication of the rights holder, restrictions on the use of the content and the time period over which the stated rights apply. If no usage terms are specified, then no specific restrictions on use of the content beyond contractual ones are being asserted.

14.6 Datatype specifications

All XML data types of NewsML-G2 are defined by the NewsML-G2 Schema - see how to get them in the [The Full Set of Specification Documents](#) section.

This section only adds for some data types User Notes, Implementation Notes or recommended Controlled Vocabularies which are not in the NewsML-G2 XML Schema version 2.23.

All data types not having such additional information are not listed here.

14.6.1 ApproximateDateTimePropType

User Note: If a start and/or end attribute exists, then the date is approximate, else it is defined precisely by the property's date. If only the approximation start date is provided the range ends with the property value; if only the approximation end date is provided the approximation range starts with the property value.

14.6.2 AudienceType

User Note:

significance: 1 – corresponds to the highest significance.

significance: 9 – corresponds to the lowest significance.

14.6.3 BlockType

User Note: Blocks are primarily used for notes, comments or instructions created by a news provider for use by recipient editorial teams.

14.6.4 ConceptIdType

User Note: Rule for @qcode and @uri in an element: If both attributes, @qcode and @uri, are present the @qcode takes precedence.

14.6.5 DateOptTimePropType and DateOptTimeType

User Note: The time may be expressed in Universal Time Coordinates (UTC), or in local time



together with a time zone offset in hours and minutes

14.6.6 Flex...PropType (multiple)

Included: Flex1PropType, FlexLocationPropType, FlexOrganisationPropType, FlexPartyPropType, FlexPersonPropType, FlexPropType, FlexProp2Type

User Note: Rule for @qcode and @uri in an element:

- An element SHOULD NOT use both a @qcode and a @uri.
- If both attributes, @qcode and @uri, are present the @qcode takes precedence

14.6.7 Label1Type

User Note:

Labels are assertions expressed as natural language strings intended to be consumed by human beings. They are typically displayed alongside the content of an Item or in place of Items in a list, providing a means of selection among them.

14.6.8 Link1Type

User Note: To identify the target resource either the residref attribute or the href attribute MUST be set, optionally both MAY be used in parallel. The residref attribute identifies the target resource by its globally unique identifier (if the resource has such an identifier), while the href attribute identifies the location of the target resource in e.g. a (remote) file system. If the target resource is an Item and the residref attribute is used, a version attribute MAY indicate the target Item version; in the absence of version information, the target resource is the latest version available.

14.6.9 RecurrenceRuleType

User Note: The different datatypes of the attributes of this data type correspond to iCalendar datatypes and enumerations.

14.6.10 TruncatedDateTimePropType and TruncatedDateTimeType

Implementation Note: TruncatedDateTimePropType is used as a property datatype.

Values may look like that:

YYYY-MM-DD"T"hh:mm:ss.sssTZ

YYYY-MM-DD"T"hh:mm:ssTZ

YYYY-MM-DD

YYYY-MM

YYYY

14.7 Attribute (Group) Specifications

All XML attributes and groups of attributes of NewsML-G2 elements are defined by the NewsML-G2 Schema - see how to get them in the [The Full Set of Specification Documents](#) section.

This section only adds for some attributes User Notes, Implementation Notes or recommended Controlled Vocabularies which are not in the NewsML-G2 XML Schema version 2.23.

All attributes not having such additional information are not listed here.

14.7.1 Internationalization Attributes - i18nAttributes

Notes:

- *xml:lang* values MUST follow RFC 4646 and RFC 4647 (as both replace RFC 3066) or its successor. See also IETF BCP47.
- The *dir* qualifier specifies the directionality of scripted text: left-to-right ("ltr", the default) or right-to-left ("rtl"). Its definition follows the XHTML 1.0 production. Directionality – left-to-right or right-to-left – is assigned to characters in Unicode, in order to allow the text to be rendered properly. For example, while English characters are presented left-to-right, Hebrew characters are presented right-to-left. Unicode defines a bidirectional algorithm that must be applied whenever a document contains right-to-left characters. While this algorithm usually gives the proper presentation, some situations leave directionally neutral text and require the *dir* attribute to specify the base directionality.

14.7.2 Ranking Attributes - rankingAttributes

Processing rules for the *rank* attribute:



Properties with a lower value of the rank attribute have a higher importance than properties with a higher value of this attribute. All properties with the same value of the rank attribute have the same importance. All properties without a rank attribute have the same importance, which is lower than the importance of all properties with this attribute.

If relative importance is being used to determine display order, then:

- Properties with a lower value of the rank attribute should be displayed before properties with a higher value of this attribute.
- Properties with the same value of the rank attribute should be ordered within this rank alphabetically by their names if these are available. If some or all of the names are available in multiple languages, the order of the properties will depend on the language chosen.
- All properties without a rank attribute should be displayed after all properties with this attribute.

Examples (using rank with the language property):

```
<!-- Rank as: all equal (implicit) -->
<language tag="en"/>
<language tag="fr"/>
<language tag="es"/>
<language tag="de"/>
<!-- Rank as: en, then any others -->
<language tag="en" rank="1"/>
<language tag="fr"/>
<language tag="es"/>
<language tag="de"/>
<!-- Rank as: en, then fr, then es, then de -->
<language tag="en" rank="1"/>
<language tag="fr" rank="2"/>
<language tag="es" rank="3"/>
<language tag="de" rank="4"/>
<!-- Rank as: en, then fr, then any others -->
<language tag="en" rank="1"/>
<language tag="fr" rank="2"/>
<language tag="es"/>
<language tag="de"/>
<!-- Rank as: en and fr, then any others -->
<language tag="en" rank="1"/>
<language tag="fr" rank="1"/>
<language tag="es"/>
<language tag="de"/>
```

14.7.3 Quantify Attributes - *quantifyAttributes*

Notes:

- An indication of confidence is usually obtained by automatic categorization means. 100 is the highest value.
- A high relevance indicates that this piece of metadata truly expresses what the piece of news is about, while a low relevance indicates a low correlation between the metadata and the essence of the piece of news.
- why indicates whether the metadata is directly extracted from the content by a tool and/or by a person, whether it is an ancestor of some other concept directly associated with the content (e.g. the concepts France and Europe are ancestors of the concept Paris), or whether it is derived by look-up in a thesaurus (e.g. the entity Merck may be associated with the concept Pharmaceutical Industry Sector).

14.7.4 Orientation Attribute - *orientation*









The table below enumerates the allowed values for the **orientation** attribute. The values are integers from 1 to 8 and reflect the TIFF 6.0 and Exif 2.3 specification. Orientation 1 is considered as default value.

Remark on the Definition column: by the Exif specification the "0th row" is the first row which has been scanned for the digital image and the "0th column" the first column. The hint describes how a picture of this orientation has to be flipped and/or rotated to show as the default orientation 1.

The column "Visual example" shows a picture of the character F having an orientation aligning with the value.

The letters T(op), L(ef), R(ight) and B(ottom) represent the visual alignment of the image with orientation 1.

Table Orientation Values

Value	Definition and Explanation	Visual Example
1	The 0th row is at the visual top of the image, and the 0th column is the visual left-hand side. Hint: no action required.	
2	The 0th row is at the visual top of the image, and the 0th column is the visual right-hand side. Hint: flip horizontal.	
3	The 0th row is at the visual bottom of the image, and the 0th column is the visual right-hand side. Hint: rotate 180 degrees.	
4	The 0th row is at the visual bottom of the image, and the 0th column is the visual left-hand side. Hint: flip horizontal and rotate 180 degrees.	
5	The 0th row is the visual left-hand side of the image, and the 0th column is the visual top. Hint: flip vertical and rotate 90 degrees clockwise.	
6	The 0th row is the visual right-hand side of the image, and the 0th column is the visual top. Hint: rotate 90 degrees counterclockwise.	
7	The 0th row is the visual right-hand side of the image, and the 0th column is the visual bottom. Hint: flip vertical and rotate 90 degrees counterclockwise.	
8	The 0th row is the visual left-hand side of the image, and the 0th column is the visual bottom. Hint: rotate 90 degrees clockwise.	

15 Glossary

Table Glossary

Term	Definition
alias	See scheme alias.
anonymous controlled vocabulary	A controlled vocabulary that is not a scheme.
catalog	A document containing information about scheme(s).
code	A character sequence which forms a member of a controlled vocabulary.
concept	Anything that one may wish to refer to, e.g. Diplomacy, Paris, the Euro, OECD, the Japanese language, the IMF, Oil, Madonna, Olympic Games. Thus concept here has a broader meaning than is usual. This is because we are dealing with the idea of Paris, rather than with Paris itself, the idea of Oil, rather than Oil itself, and so on. Concepts fall in two broad categories: named entity and generic (or abstract) concepts. A concept may be defined by a ConceptItem.
ConceptItem	A specialised data structure containing data representing a concept. An identifier for the concept is mandatory and it may, optionally, provide information such as name, definition, relationships, etc. A concept defined by a ConceptItem is identified by a {scheme alias, code} pair. The reverse relationship does not necessarily hold. In other words, there is no requirement that each {scheme alias, code} pair has a corresponding ConceptItem. See also: representation of a ConceptItem.
concept type	A concept type allows the logical grouping of all similar concept(s), regardless of the scheme(s) the concepts belong to. Examples of concept type might be: Person, Organisation, Language, Business Sector, News Subject or Geography. A concept type is itself a concept and, as such, is represented by a code in a scheme.
concept URI	A URI which identifies a concept. A concept URI is obtained by appending the code representing this concept to the scheme URI corresponding to the scheme to which the code belongs. An abbreviated notation of a concept URI is a Qualified code, QCode.
conformance level	A layer of functionality defined by a standard. The News Architecture power conformance level is a superset of the News Architecture core conformance level, both in terms of structure and processing.
controlled vocabulary	A set of code(s), managed by some authority (e.g. a person or an organisation), employing some mechanism (e.g. an XML Schema, a Web page, an RFC, or KnowledgeItem) to maintain this set. A controlled vocabulary is either a scheme or is anonymous (i.e. an anonymous controlled vocabulary). Each code in a controlled vocabulary represents a concept.
constrained metadata container	A metadata container which either accepts only code(s) of a specified concept type or accepts only codes from a specified controlled vocabulary (which may be an anonymous controlled vocabulary or a scheme).
Definition	A human-readable string, held within a ConceptItem, which defines the concept which the item represents. Definitions will be implemented using free-form text.
formal metadata element	A metadata element designed to hold data that is not free-form text, e.g. code(s), or formal text. Such data is usually consumed by software. An example of such an element with a code value is subject. An example value of subject is "nc:15062000".
free-form metadata element	A metadata element designed to hold free-form text. Such data is usually consumed by humans. An example of a free-form metadata element is title. An example value of title is "Ian Thorpe makes a splash". The News Architecture provides a couple of datatypes for free-form text, e.g. International String, Label or BlockText.

Table 289. Glossary (Continued)



Term	Definition
free-form text	Arbitrary text, i.e. text which does not consist of code(s) drawn from a controlled vocabulary. A headline or a description is an example of free-form text.
formal text	A set of one or more metadata container(s) for free-form text to express formal information about a specific concept, but without identifying it. Basic properties for formal text are literal, name, definition and note. An example for formal text is the Creator property with a value of name = "Alfred Hitchcock", definition = "Suspense movie director and producer, born 1899, died 1980".
globally unique identifier	An identifier that is unique, unambiguous, and persistent. Being unique and unambiguous means that there is a 1:1 relationship between the identifier and the identified object. Being persistent means that the identifier never changes as time passes, and that it is never reused as an identifier for another object even if the original object disappears. See also persistent identifier, unambiguous identifier, and unique identifier.
Identifier	A string used to identify a specific resource. See persistent identifier, unambiguous identifier, unique identifier, and globally unique identifier (GUID).
KnowledgeItem	A Knowledge Item is a set of concept definitions to form a consistent structure, which is managed, protected and published as a whole. It facilitates the management and exchange of controlled vocabulary(ies).
Label	A generic term for datatypes designed to hold free-form text.
Metadata	Data which asserts something about some other data.
metadata container	A location (e.g. an element or an attribute) in a data structure, designed to hold Metadata. In XML it may be implemented as a metadata element.
metadata element	An XML element, which is either a formal metadata element or a free-form metadata element, it implements the notion of a metadata container.
named entity	A named entity may be a person, place, event, organization, product name, object name or any other news-related real life entity.
News Architecture	A framework of specifications common to all IPTC news exchange standards of the G2 Family of Standards.
news provider	A provider of news content, the entity responsible for the management of news items May be a news agency, a syndication company, a newspaper, a magazine ... or a blogger.
ontology	See taxonomy.
persistent identifier	An identifier which is associated with the same resource for all time. See also unambiguous identifier, unique identifier, and globally unique identifier (GUID).
processor	An application that supports the handling and processing of Items. Also known as a user agent.
property	A synonym term for a metadata container – may be implemented as XML element.
provider	See news provider.
publish	Make available to other parties involved in the news exchange process, according to the business practices of the provider.



Table 289. Glossary (Continued)

Term	Definition
Qualified code, QCode	A concept URI represented by a string of the form sss:ccc, where sss is a scheme alias and ccc is a code. Examples are iso4217:USD, rfc3066:zh-Hant, nc:15062000, nasdaq:msft and cusip:594918104. A QCode is not the same as a QName (qualified name) [W3C: Namespaces in XML (http://www.w3.org/TR/REC-xml-names/)], though there are substantial similarities. The two main differences are: (i) the code does not have to be a valid XML name (e.g. can start with a digit), and (ii) the scheme alias does not have to be declared using a namespace declaration.
representation	The physical form of something.
representation of a ConceptItem	A manifestation of a given ConceptItem that is suited for some particular purpose. The various representations of a given ConceptItem may differ, for example, in whether they are verbose or concise, or in which language(s) they use for name and definition.
resource	A resource is a set of data that has identity.
scheme	A controlled vocabulary which is identified by a scheme URI. A scheme is not an anonymous controlled vocabulary.
scheme alias	A character sequence which is used as an abbreviation for a scheme URI. A scheme alias is similar but not identical to an XML Namespace prefix.
scheme URI	The URI which identifies the scheme. It is recommended to make this URI a URL and resolving it should result in retrieving information about the scheme.
synonym	Synonyms are concept URI(s) that refer from one concept to another concept with equivalent semantics. Synonymy is a symmetric relationship, which means that if A is synonymous with B, then B is also synonymous with A. An example of synonyms is "cemetery" and "graveyard". In the News Architecture synonyms are expressed by the sameAs {Relationship} property.
target	The data being described by the metadata. The IPTC has chosen to use the term target rather than subject (the term used by RDF [http://www.w3.org/RDF/]), as subject has a special meaning in the context of News.
taxonomy	In a broad sense, taxonomy is the science of classification, but is often taken to mean a particular classification. In the context of the News Architecture, a taxonomy is a collection of concept(s), with associated code(s). A taxonomy may support typed relationships between concepts. Such a taxonomy is sometimes known as an ontology or thesaurus.
thesaurus	See taxonomy.
tuple	A set of values. The word tuple is a generalisation of the sequence: couple, triple, quadruple, quintuple, sextuple, etc. Tuples are conventionally written as a comma-separated list of items, enclosed within braces, e.g. {scheme alias, code}.
type	See concept type.
unambiguous identifier	An identifier is unambiguous if it identifies one and only one object (but an object may have several different unambiguous identifiers). See also globally unique identifier.
unconstrained metadata	A metadata container that accepts code(s) from any controlled vocabulary and of any concept type.
unique identifier	The only identifier of a resource. See also persistent identifier, unambiguous identifier, and globally unique identifier (GUID)
Web resource	The data content that can be retrieved from a Web server using a Web-compliant transport protocol. See also resource.



16 References

Subject	Description
IPTC Documents	
NML-BR	IPTC NewsML 2 Business Requirements: http://www.iptc.org/std/NewsML/2.0/specification/NewsML_2.0-spec-BusinessRequirements_1.pdf
EventsML-G2	Specifications for EventsML-G2: http://www.iptc.org/std/NewsML-G2/2.23/specification/
NewsML-G2	Specifications for NewsML-G2: http://www.iptc.org/std/NewsML-G2/2.23/specification/
IPTC NewsCodes	All IPTC codes to categorise content or to express functional features can be obtained as NewsCodes from: http://www.newscodes.org
Other References	
RFC2119	Key words for use in RFCs to Indicate Requirement Levels http://www.ietf.org/rfc/rfc2119.txt
XMLSCHEMA-1.0 XSD	W3C XML Schema 1.0 specifications at: http://www.w3.org/XML/Schema
XMLDSIG	XML-Signature Syntax and Processing: http://www.w3.org/TR/xmlsig-core/
RDF	Resource Description Framework (RDF): http://www.w3.org/RDF/
BCP47	Tags for Identifying Languages, IETF: http://www.rfc-editor.org/rfc/bcp/bcp47.txt
iCalendar	iCalendar as specified by RFC 2445: http://www.ietf.org/rfc/rfc2445.txt

Contact Information

Contact the IPTC by:

Web: www.iptc.org

Twitter: @IPTC and @IPTCupdates

Email: office@iptc.org

Postal mail:

25 Southampton Buildings

London WC2A 1AL United

Kingdom