IPTC Standards

NewsML™ G2

(version 2.21)

Quick Start Guide to NewsML-G2 Basics

Quick Start: NewsML-G2 Basics

About the Quick Start Guides

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

Introduction

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

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

Item structure

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

Basic data structure of any NewsML-G2 Item

- Catalog
- Rights metadata
- Metadata about the Item as a whole
- Metadata about all or part of the Item content
- Helper structures
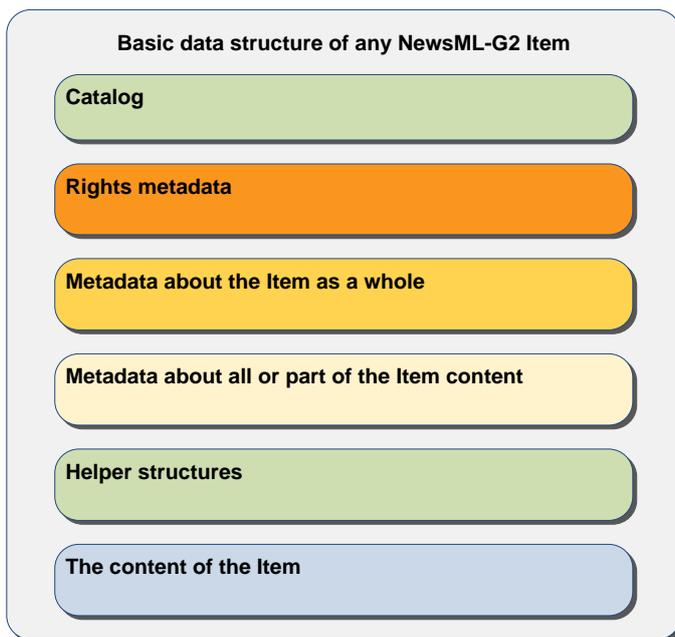- The content of the Item

Figure 1: All G2 Items share this basic container structure

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

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

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

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

Optional "helper" structures are available for specialised processing needs; their use is covered in **Chapter 25 Advanced Metadata Techniques** in the main Guidelines document.

Each type of G2 Item has a specific wrapper element for content, shown in the diagram below that also shows the basic top level elements common to all G2 Items:
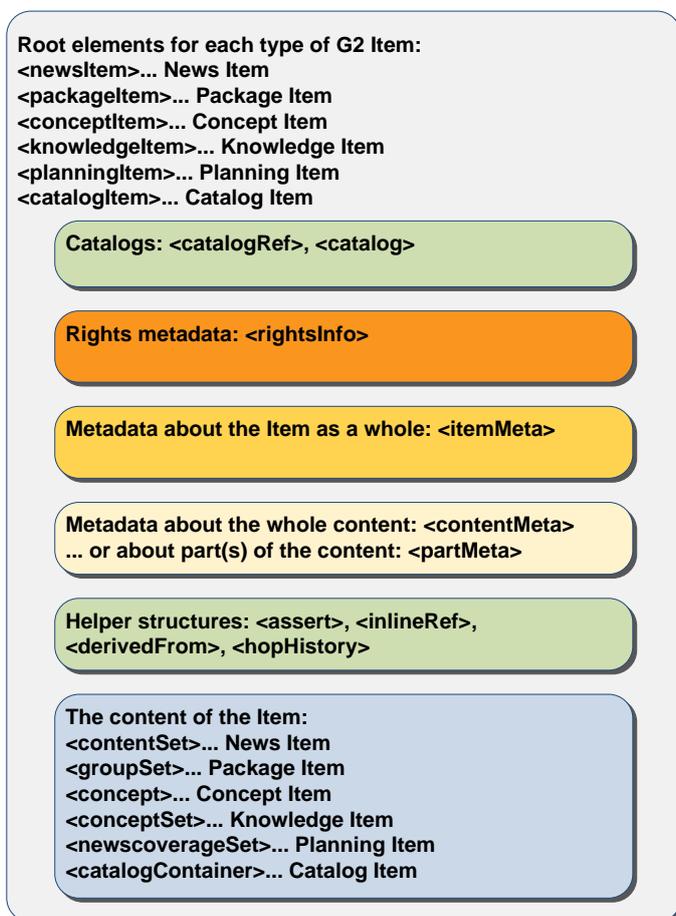
Root elements for each type of G2 Item:
<newsItem>... News Item
<packageItem>... Package Item
<conceptItem>... Concept Item
<knowledgeItem>... Knowledge Item
<planningItem>... Planning Item
<catalogItem>... Catalog Item

Catalogs: <catalogRef>, <catalog>

Rights metadata: <rightsInfo>

Metadata about the Item as a whole: <itemMeta>

Metadata about the whole content: <contentMeta>
... or about part(s) of the content: <partMeta>

**Helper structures: <assert>, <inlineRef>,
<derivedFrom>, <hopHistory>**

The content of the Item:
<contentSet>... News Item
<groupSet>... Package Item
<concept>... Concept Item
<conceptSet>... Knowledge Item
<newscoverageSet>... Planning Item
<catalogContainer>... Catalog Item

Figure 2: The basic XML elements associated with each part of a G2 Item

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

"Invisible" values

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

Root element

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

Root element attributes

Item Identifier

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

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

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

```
<newsItem guid="urn:newsml:acmenews.com:20151018:US-FINANCE-FED"
```

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

Version

A simple indicator of the version of the Item:

```
version="8"
```

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

Standard

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

```
standard="NewsML-G2"
```

Standard Version

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

```
standardversion="2.21"
```

Conformance

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

```
conformance="power"
```

Language

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

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

IPTC namespace

Sets the default namespace for elements:

```
xmlns="http://iptc.org/std/nar/2006-10-01/"
```

Putting this together, the required root element attributes are:

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:acmenews.com:20151018:US-FINANCE-FED"
  version="8"
  standard="NewsML-G2"
  standardversion="2.21"
  conformance="power"
  xml:lang="en-GB">
```

Validating code

When developing a NewsML-G2 processing application, implementers will need to validate the generated G2 code against the appropriate schema. To validate code at PCL against the latest version of the NewsML-G2 specification (2.21) add the following code to the <newsItem> element:

```
xmlns="http://iptc.org/std/nar/2006-10-01/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://iptc.org/std/nar/2006-10-01/
    http://www.iptc.org/std/NewsML-G2/2.21/specification/NewsML-G2_2.21-spec-All-
Power.xsd"
```

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

`NewsML-G2_2.21-spec-All-Core.xsd`

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

Catalogs

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

One of the few mandatory NewsML-G2 elements, <itemClass>, uses QCodes issued by the IPTC to identify the business intention of the Item. For a News Item, the scheme is News Item Nature, with a recommended alias of "ninat". Values from the scheme include (not limited to) "ninat:text" and "ninat:picture". The catalog reference is:

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

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

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

```
<catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_27.xml" />
```

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

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

```
<catalogRef
    href="http://www.example.com/newsml-g2/catalog.enews_2.xml" />
```

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

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
    guid="urn:newsml:acmenews.com:20151018:US-FINANCE-FED"
    version="8"
    standard="NewsML-G2"
    standardversion="2.21"
    conformance="power"
    xml:lang="en-GB">
    <catalogRef
        href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_27.xml" />
    <catalogRef
        href="http://www.example.com/newsml-g2/catalog.enews_2.xml" />
```

Rights information

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

```
<rightsInfo>
    <copyrightHolder uri="http://www.example.com/about.html#copyright" >
```

```
            <name>Example Enews LLP</name>
        </copyrightHolder>
        <copyrightNotice>
            Copyright 2015 Example Enews LLP, all rights reserved
        </copyrightNotice>
        <usageTerms>
            Not for use outside the United States
        </usageTerms>
    </rightsInfo>
```

Item Metadata <itemMeta>

Mandatory Properties

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

<u>Item Class</u>

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

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

Other possible values from this scheme include (not limited to) "ninat:picture", "ninat:video" and "ninat:audio".

<u>Provider</u>

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

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

<u>Version Created</u>

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

```
<versionCreated>2015-10-21T16:25:32-05:00</versionCreated>
```

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

<u>Publication Status</u>

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

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

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

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

When an Item is cancelled, it must never be used, but an Item that has been withheld may subsequently have its status updated to "usable". For further information on processing rules, see **section 26.2** of the main Guidelines.

Use of URIs in place of QCodes

The original NewsML-G2 developers wanted to use URIs as the preferred way to identify concepts, as this would enable the resulting concept identifiers to be globally unique and optionally to be a reference to a

web resource. The constraints on network bandwidth (compared to today) led the developers to propose QCodes to represent URIs, because:

- ❖ they are lightweight and economise file sizes,
- ❖ although they are designed to reference web resources, the delivery of data in response to requesting QCodes that are resolved to full URIs is optional; the codes may be used "as is" without de-referencing to the full URI.

Later, some providers asked for flexibility to use full URIs in property values, so from NewsML-G2 v2.11 this feature was added enabling @uri to be used in place of, or in parallel with, @qcode. (where both are used, @qcode takes precedence). The <pubStatus> assertion:

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

could be expressed using a URI as:

```
<pubStatus uri="http://cv.iptc.org/newscodes/pubstatusg2/usable" />
```

Further, in NewsML-G2 v2.18 onwards, other properties with QCode Type values had "URI siblings" added, for example properties with @role may now have a value expressed as @roleuri. In NewsML-G2 v2.20, properties with a mandatory @qcode were changed so that @uri may be used instead. For details of these changes, see **Chapter 14** of the full Guidelines.

Optional Properties

The following optional properties are frequently used by G2 providers.

First Created

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

```
<firstCreated>2015-10-18T13:12:21-05:00</firstCreated>
```

Embargoed

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

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

```
<embargoed>2015-10-23T12:00:00Z</embargoed>
```

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

Service

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

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

Editorial Note

The <edNote> element contains a note that is intended to be read by internal staff at the receiving organisation, but not published to the end-user; in this example it conveys some optional information about the release condition imposed by the <embargoed> element:

```
<edNote>
    Note to editors: STRICTLY EMBARGOED. Not for public release until 12noon
    on Friday, October 23, 2015.
</edNote>
```

Signal

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

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

The other value in the Signal scheme is "correction". There is further information on giving fine-grained information about signalling the reason and impact of updates in the **section 26.6 Processing Updates and Corrections** in the main Guidelines document. For guidance on applying signals to different parts of an Item's content, see **Chapter 25.8** of the main Guidelines.

Link

The <link> element has two basic purposes:

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

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

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

Completed Item Metadata

```
<itemMeta>
    <itemClass qcode="ninat:text" />
    <provider qcode="nprov:REUTERS" />
    <versionCreated>2015-10-21T16:25:32-05:00</versionCreated>
    <firstCreated>2015-10-18T13:12:21-05:00</firstCreated>
    <embargoed>2015-10-23T12:00:00Z</embargoed>
    <pubStatus qcode="stat:usable" />
    <service qcode="svc:uknews">
        <name>UK News Service</name>
    </service>
    <edNote>
        Note to editors: STRICTLY EMBARGOED. Not for public release until 12noon
        on Friday, October 23, 2015.
    </edNote>
    <signal qcode="sig:update" />
    <link rel="irel:seeAlso"
        href="http://www.example.com/video/20081222-PNN-1517-407624/index.html"/>
</itemMeta>
```

Content Metadata <contentMeta>

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

Administrative Metadata

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

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

Timestamps

The <contentCreated> timestamp corresponds to a "Created on" field of the story. It is expressed in NewsML-G2 as Truncated DateTime data type, meaning that the date-time elements may optionally be

stripped, starting from the right. If required, the <contentModified> property may also be used to contain the "Last Edit" timestamp. This must be later than the Created timestamp.

```
<contentCreated>2015-10-21T15:21:06-05:00</contentCreated>
<contentModified>2015-10-21T16:22:45-05:00</contentModified>
```

Located

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

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

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

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

Broader, Narrower

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

```
<located type="cptype:city" qcode="geo:345678">
    <name>Berlin</name>
    <broader type="cptype:statprov" qcode="prov:2365">
        <name>Berlin</name>
    </broader>
    <broader type="cptype:country" qcode="iso3166-1a2:DE">
        <name>Germany</name>
    </broader>
</located>
```

Creator

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

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

Information Source

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

```
<infoSource uri="http://www.example.com/pressreleases/201510/newproducts.html" />
```

Descriptive Metadata

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

Subject

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

```
<subject type="cpnat:abstract" qcode="medtop:20000523" />
<subject uri="http://cv.iptc.org/newscodes/mediatopic/20000523" />
```

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

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

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

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

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

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

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

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

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

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

Genre

The <genre> element indicates the style of the content, in this example "interview" as a property that is distinct from <subject> that is used to indicate the subject matter of content. In the example, an IPTC Genre NewsCodes value of "Interview" is used:

```
<genre qcode="genre:interview">
```

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

Slugline

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

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

Headline

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

```
        <headline>Fed to halt QE to avert "bubble"</headline>
```

Completed Content Metadata

```
    <contentMeta>
        <contentCreated>2015-10-21T15:21:06-05:00</contentCreated>
        <contentModified>2015-10-21T16:22:45-05:00</contentModified>
        <located type="cptype:city" qcode="geo:345678">
            <name>Berlin</name>
            <broader type="cptype:statprov" qcode="prov:2365">
              <name>Berlin</name>
            </broader>
            <broader type="cptype:country" qcode="iso3166-1a2:DE">
              <name>Germany</name>
            </broader>
        </located>
        <creator uri="http://www.example.com/staff/mjameson" >
            <name>Meredith Jameson</name>
        </creator>
        <infoSource uri="http://www.example.com" />
        <subject type="cpnat:abstract" qcode="medtop:04000000">
            <name xml:lang="en-GB">economy, business and finance</name>
        </subject>
        <subject type="cpnat:abstract" qcode="medtop:20000523">
            <name xml:lang="en-GB">labour market</name>
            <name xml:lang="de">Arbeitsmarkt</name>
            <broader qcode="medtop:04000000" />
        </subject>
        <genre qcode="genre:interview">
            <name xml:lang="en-GB">Interview</name>
        </genre>
        <slugline>US-Finance-Fed</slugline>
        <headline>Fed to halt QE to avert "bubble"</headline>
    </contentMeta>
```

Content

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

Code Listing: A NewsML-G2 News Item

All Scheme Aliases used in listing below indicate IPTC NewsCodes vocabularies, except for the following alias values: *svc, cptype, geo, prov.*

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:acmenews.com:20151021:US-FINANCE-FED"
  version="8"
  standard="NewsML-G2"
  standardversion="2.21"
  conformance="power"
  xml:lang="en-GB">
  <catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_27.xml" />
```

```xml
    <catalogRef href="http://www.example.com/newsml-g2/catalog.enews_2.xml" />
    <rightsInfo>
        <copyrightHolder uri="http://www.example.com/about.html#copyright" >
            <name>Example Enews LLP</name>
        </copyrightHolder>
        <copyrightNotice>
            Copyright 2015 Example Enews LLP, all rights reserved
        </copyrightNotice>
        <usageTerms>
            Not for use outside the United States
        </usageTerms>
    </rightsInfo>
    <itemMeta>
        <itemClass qcode="ninat:text" />
        <provider qcode="nprov:REUTERS" />
        <versionCreated>2015-10-21T16:25:32-05:00</versionCreated>
        <firstCreated>2015-10-18T13:12:21-05:00</firstCreated>
        <embargoed>2015-10-23T12:00:00Z</embargoed>
        <pubStatus qcode="stat:usable" />
        <service qcode="svc:uknews">
            <name>UK News Service</name>
        </service>
        <edNote>
            Note to editors: STRICTLY EMBARGOED. Not for public release until 12noon
            on Friday, October 23, 2015.
        </edNote>
        <signal qcode="sig:update" />
        <link rel="irel:seeAlso"
            href="http://www.example.com/video/20081222-PNN-1517-407624/index.html"/>
    </itemMeta>
    <contentMeta>
        <contentCreated>2015-10-21T15:21:06-05:00</contentCreated>
        <contentModified>2015-10-21T16:22:45-05:00</contentModified>
        <located type="cptype:city" qcode="geo:345678">
            <name>Berlin</name>
            <broader type="cptype:statprov" qcode="prov:2365">
                <name>Berlin</name>
            </broader>
            <broader type="cptype:country" qcode="iso3166-1a2:DE">
                <name>Germany</name>
            </broader>
        </located>
        <creator uri="http://www.example.com/staff/mjameson" >
            <name>Meredith Jameson</name>
        </creator>
        <infoSource uri="http://www.example.com" />
        <subject type="cpnat:abstract" qcode="medtop:04000000">
            <name xml:lang="en-GB">economy, business and finance</name>
        </subject>
        <subject type="cpnat:abstract" qcode="medtop:20000523">
            <name xml:lang="en-GB">labour market</name>
            <name xml:lang="de">Arbeitsmarkt</name>
            <broader qcode="medtop:04000000" />
        </subject>
        <genre qcode="genre:interview">
            <name xml:lang="en-GB">Interview</name>
        </genre>
        <slugline>US-Finance-Fed</slugline>
        <headline>Fed to halt QE to avert "bubble"</headline>
    </contentMeta>
    <contentSet>
        <inlineXML contenttype="application/nitf+xml"> <!--  A VALID MIME TYPE    -->
            <!--       Inline XML must contain well-formed XML such as XHTML       -->
        </inlineXML>
    </contentSet>
</newsItem>
```

News Content options for News Items and Package Items

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

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

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

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

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

Summary and Next Steps

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

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

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

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

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

Hidden Values of NewsML-G2

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

All G2 items

- ❖ @version of the root element = "1"
- ❖ @conformance of the root element = "core"
- ❖ <embargoed> = no embargo
- ❖ <pubStatus> = "usable"
- ❖ <catalog> and <catalogRef> = one is required; many of either, or a mix of both may be used

❖ @scope of <hash> = "content". Hash value is a message digest included in the Item for security purposes. A hash scope of "content" indicates that the hash value was derived by hashing some/all of the content only.

❖ @why (attribute of many elements) = "direct". The attribute value indicates that the value is directly related to the content.

❖ @how (attribute of many elements) = "person". The attribute indicates how the value was extracted from the content: by a person. (See **Why and How metadata has been added: @why and @how** in the main Guidelines for essential guidance on the use of @why and @how.)

❖ @custom (attribute of many elements) = "false". The attribute indicates that the property was added specifically for a customer or group of customers

❖ @dir (many elements) = "ltr". The directionality of the script of the language of the property is left to right.

News Items only

❖ @timeunit (when @duration is used) = "seconds"

❖ @dimensionunit (when @width/@height is used) = (for example) pixels for a still picture. See the Quick Start Guide – Pictures for details.

❖ @encoding of <inlineData> = "base64". This is the default binary encoding for Inline Data.

Package Item only

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

NewsML-G2 Implementation Guidelines and Specification

For more comprehensive information about G2 implementation than is covered by these Quick Start Guides, the full Guidelines for NewsML-G2 Implementers may be downloaded from www.newsml-g2.org/doc.

This includes more detailed "How To" topics to help implementers with more complex needs, and also covers subjects such as creating and managing Catalogs and Controlled Vocabularies, conveying multiple G2 Items in News Messages, and using Planning and EventsML-G2 for news management and fulfilment.

The NewsML-G2 Specification is available for download at www.newsml-g2.org/spec.

Terms of Use

Copyright © 2015 IPTC, the International Press Telecommunications Council. All Rights Reserved.

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

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

If you have any questions about the terms, please contact the managing director of the International Press Telecommunication Council. Contact details of the IPTC are listed below.

While every care has been taken in creating this document, it is not warranted to be error-free, and is subject to change without notice. Check for the latest version of this Document and applicable G2 Standards and Documentation by visiting www.newsml-g2.org/doc. The version of NewsML-G2 covered by this document is 2.21.

Contacting the IPTC

IPTC, International Press Telecommunications Council
Web address: www.iptc.org
Follow us on Twitter: @IPTC and @IPTCupdates
Email: office@iptc.org
Business address
25 Southampton Buildings
London WC2A 1AL
United Kingdom

The company is registered in England at 10 Portland Business Centre, Datchet, Slough, Berks, SL3 9EG
as Comité International des Télécommunications de Presse
Registration No. 1010968, Limited by Guarantee, Not Registered for VAT