# IPTC Standards

NewsML™ G2

(version 2.23)

# Quick Start Guide to Receiving NewsML-G2

Public Release

Document Revision 1.0.1

# Quick Start: Receiving NewsML-G2

This document is designed to be usable by a NewsML-G2 provider as a template for a guideline to their customers, detailing the NewsML-G2 that is expected to be sent from that specific provider. It uses the virtual company name NAG (News Agency), which should be replaced by the name of the provider using this document. These parts are indicated by this orange text. It is also intended that the provider can customise the document, deleting what is not needed and adding their own details as required.

It is recommended that the receiver reads the IPTC's Quick Start Guide to NewsML-G2 Basics before this Quick Start Guide, as it contains helpful information about the structure of NewsML-G2.

## Introduction

NewsML-G2 is one of the latest of the IPTC news exchange standards, building on the experience and legacy of earlier standards such as IPTC7901, IIM and NewsML 1, which are still in use. As such, there are a number of common Use Cases that any news receiver needs to support. This Quick Start Guide addresses the basics of processing an incoming NewsML-G2 News Item based on the most common properties and specific properties as they are used by NAG.

## Example

Below is a view of an example story and supporting information received from NAG as it might be displayed to a journalist working for a customer organisation:

| NAG Customer – Content Editing System | | |
|---|---|---|
| Slugline | US-Finance-Fed | |
| Created on | 2016-11-21 15:21:06 | |
| Source | NAG | |
| Author | mjameson | |
| Latest edit | 2016-11-21 16:22:45 | |
| Latest editor | moiras | |
| Categories | economy, finance, business, central bank, monetary policy | |
| Headline | Fed to halt QE to avert "bubble" | |
| Byline | By Meredith Jameson | |
| Dateline | Washington D.C. | 21/11/2016 |
| Body Text | Et, sent luptat luptat, commy nim zzriureet vendreetue modo dolenis ex euisis nosto et lan ullandit lum doloreet vulla feugiam coreet, cons eleniam il ute facin veril et aliquis ad minis et lor sum del iriure dit la feugiamcommy nostrud min ullapat velisl duisismodip ero dipit nit utpatum sandrer cipisim nit lortis augiat nulla faccum at am, quam velenis nulput la auguerostrud magna commolore eliquatie exerate facilis modiamconsed dion henisse quipit at. Ut la feu facilla feu faccumsan ecte modoloreet ad ex el utat.<br>. | |

The original NewsML-G2 document that was used to generate the above view is shown below:

### Code Listing: NewsML-G2 Text Document

(All Scheme Aliases used in the listing below indicate IPTC NewsCodes vocabularies, except for the following alias values: *geoloc, is*)

Each wrapper element is highlighted using a different background colour to aid identification. A logical diagram of this structure is shown in the I{PTC Quick Start Guide to NewsML-G2 Basics.

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:NAG.com:20161121:US-FINANCE-FED"
  version="9"
  standard="NewsML-G2"
  standardversion="2.23"
  xml:lang="en-US">
```

```xml
<catalogRef
    href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_29.xml" />
<catalogRef
    href="http://catalog.NAG.com/news/NAG_G2_CODES_2.xml" />
<rightsInfo>
    <copyrightHolder uri="http://www.NAG.com/about.html#copyright">
        <name>NAG</name>
    </copyrightHolder>
    <copyrightNotice>Copyright 2016 NAG</copyrightNotice>
    <usageTerms>NAG Example: Contact your local office for all commercial or
        promotional uses. Full editorial rights UK, US, Ireland, Canada (not
        Quebec). Restricted editorial rights for daily newspapers elsewhere,
        please call.</usageTerms>
</rightsInfo>
<itemMeta>
    <itemClass qcode="ninat:text" />
    <provider uri="http://www.NAG.com/about/" />
    <versionCreated>2016-11-21T16:25:32-05:00</versionCreated>
    <pubStatus qcode="stat:usable" />
</itemMeta>
<contentMeta>
    <urgency>5</urgency>
    <contentCreated>2016-11-21T15:21:06-05:00</contentCreated>
    <contentModified>2016-11-21T16:22:45-05:00</contentModified>
    <located qcode="geoloc:NYC">
        <name>New York, NY</name>
    </located>
    <creator uri="http://www.NAG.com/staff/mjameson">
        <name>Meredith Jameson</name>
    </creator>
    <infoSource qcode="is:AP">
        <name>Associated Press</name>
    </infoSource>
    <language tag="en-US" />
    <subject qcode="medtop:04000000">
        <name>economy, business and finance</name>
    </subject>
    <subject qcode="medtop:20000350">
        <name>central bank</name>
    </subject>
    <subject qcode="medtop:20000379">
        <name>money and monetary policy</name>
    </subject>
    <slugline>US-Finance-Fed</slugline>
    <headline>Fed to halt QE to avert "bubble"</headline>
</contentMeta>
<contentSet>
    <inlineXML contenttype="application/nitf+xml">
        <nitf xmlns="http://iptc.org/std/NITF/2006-10-18/">
            <body>
                <body.head>
                    <hedline>
                        <hl1>Fed to halt QE to avert "bubble"</hl1>
                    </hedline>
                    <byline>By Meredith Jameson, <byttl>NAG Reporter</byttl></byline>
                </body.head>
                <body.content>
                    <p>(New York, NY - November 21) Et, sent luptat luptat, commy Nim
                        zzriureet vendreetue modo dolenis ex euisis nosto et lan ullandit
                        lum doloreet vulla.</p>
                    <p>Ugiating ea feugait utat, venim velent nim quis nulluptat num
                        Volorem inci enim dolobor eetuer ercin utpatio dolorpercing.</p>
                </body.content>
            </body>
        </nitf>
    </inlineXML>
</contentSet>
</newsItem>
```

✋ NAG wraps one or more NewsML-G2 Items in a News Message (<newsMessage>), an optional XML component for managing the transmission and reception of Items. See Exchanging News: News Messages in the IPTC NewsML-G2 Guidelines for details.

At a minimum, a News Message has a <header> with a <timestamp> child element, and an <itemSet> containing one or more NewsML-G2 Items. Apart from recording the receipt of Items, the properties of a News Message are intended to be transient, and are NOT inherited by the editorial workflow.

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

## How the Item is identified

The top level (root) element <newsItem> uniquely identifies this NewsML-G2 document using @guid and @version:

```
<newsItem xmlns="http://iptc.org/std/nar/2006-10-01/"
  guid="urn:newsml:NAG.com:20161121:US-FINANCE-FED"
  version="9"
  standard="NewsML-G2"
  standardversion="2.23"
  conformance="power"
  xml:lang="en-US">
```

## Catalogs and QCodes

This lists references to the Catalogs used to resolve QCodes in the Item:

```
<catalogRef
  href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_29.xml" />
<catalogRef
  href="http://catalog.NAG.com/news/NAG_G2_CODES_2.xml" />
```

The processing of QCodes is mentioned as an area of uncertainty in feedback from some receivers of NewsML-G2. At the simplest level, QCodes are a method of guaranteeing the precision of metadata values across many documents and sources. For example a provider uses a code "HHH" as a category of news. Another provider may use "HHH" to mean a different category. If the code is conveyed in a QCode format, the Catalog enables the receiver to differentiate between these two seemingly identical values by expanding each value into a globally-unique URI (termed a "Concept URI").

So it may be stated that there are essentially two approaches to processing QCodes:

- ❖ Use the values "as is" – in other words in the same way that such a code would have been processed in IPTC7901
- ❖ Resolve the QCode either partially using the information in the Catalog to reference a unique Concept URI, or fully by retrieving the full information from the Concept URI (if available)

Further information about resolving QCodes is covered in depth in the IPTC NewsML-G2 Guidelines; for the purposes of this Quick Start Guide, we will use the QCodes "as is". Note also that it is possible to replace a QCode with a full URI, as in:

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

See the Chapter on Controlled Vocabularies in the IPTC NewsML-G2 Guidelines and the section How QCodes Work.

## Rights to use the Content

The <rightsInfo> block contains copyright information and may include a block of natural language text in <usageTerms> giving details of any rights restrictions.

```
<rightsInfo>
  <copyrightHolder uri="http://www.NAG.com/about.html#copyright" >
    <name>NAG</name>
  </copyrightHolder>
  <copyrightNotice>Copyright 2016 NAG</copyrightNotice>
  <usageTerms>NAG Example: Contact your local office for all commercial or
    promotional uses. Full editorial rights UK, US, Ireland, Canada (not
    Quebec). Restricted editorial rights for daily newspapers elsewhere,
    please call.</usageTerms>
</rightsInfo>
```

## News Management

The <itemMeta> wrapper contains information that is essential to an editorial workflow

## Can you publish this Item?

Three elements define the usability of the Item and its contents:

- ❖ pubStatus
- ❖ embargoed
- ❖ signal

```
<itemMeta>
    …
    <pubStatus qcode="stat:usable" />
    <embargoed>2016-10-23T12:00:00Z</embargoed>
    <edNote>
        Note to editors: STRICTLY EMBARGOED. Not for release until 12noon GMT on
        October 23.
    </edNote>
    <signal qcode="sig:update" />
</itemMeta>
```

Although it is recommended practice to make the publication status explicit ( as in the example) <pubStatus> may be omitted. If so, the default status is "usable". Other values are "withheld", indicating that the Item may become usable (again) at some point in the future, or "canceled", meaning the Item and all versions of it must be deleted from all systems (including archives). Note the U.S. spelling of usability terms. (See Publishing Status in the IPTC NewsML-G2 Guidelines)

The <embargoed> element indicates whether there is any time restriction affecting publication. The publication status will be "usable", but the content of the Item may not be published until the embargo time has passed. (See Embargo in the IPTC NewsML-G2 Guidelines)

The <signal> element is used to tell receivers about special ways of processing this item; this may include:

- ❖ further information about the reasons for a new version of an Item being sent. Valid values are the self-explanatory "update" and "correction".
- ❖ (NAG-specific uses of signal) …

See Processing Updates and Corrections in the IPTC NewsML-G2 Guidelines.

## Editorial Notes

The <edNote> element carries a natural-language note or instruction addressed to receivers and their staff, which MAY NOT be intended to be seen by end-users, since it may contain privileged or sensitive information.

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

The rules for publishing content of an Editorial Note are:

- ❖ If the Editorial Note has a pubconstraint attribute value indicating that the information must NOT be published the receiver must comply.
- ❖ If the pubconstraint value indicates that the information MAY be published it is the receiver's decision whether to publish or not.

Typically, background information about corrections, the withdrawal or cancellation of an item; information that is intended only for communication to editorial staff, would NOT be published; information about contacts related to the content may be published according to the editors' choice.

## Type of Content in the Item

The classification of the content <itemClass> is a mandatory property in the Item Metadata wrapper. In this case, the content is "text". Values include "picture", "video" and "composite". :

```
<itemMeta>
    <itemClass qcode="ninat:text" />
    …<signal qcode="sig:update" />
```

The controlled values for Item Class are maintained by the IPTC as part of the NewsCodes, a standard set of codes intended to promote inter-operability. The NewsML-G2 specification mandates the use of the

News Item Nature NewsCodes for the Item Class property. Visit http://iptc.org/std/newscodes/ninature to view this controlled vocabulary in full.

## Timestamps

The timestamp of this version of the Item is contained in <versionCreated>:

```
<itemMeta>
    …
    <versionCreated>2016-11-21T16:25:32-05:00</versionCreated>
    …
```

Content Timestamps are supported in the Content Metadata wrapper <contentMeta>. The timestamp of the first / current version of the Content is contained in <contentCreated> / <contentModified> respectively:

```
<contentMeta>
    …
    <contentCreated>2016-11-21T15:21:06-05:00</contentCreated>
    <contentModified>2016-11-21T16:22:45-05:00</contentModified>
    …
```

See Timestamps in IPTC NewsML-G2 Guidelines for more information about date-time processing, but note that <versionCreated> is the only mandatory timestamp in NewsML-G2 and is the default value for other timestamp properties if they are omitted.

NAG-specific rules for the use of timestamps are: …

## Content Management

The <contentMeta> wrapper contains metadata about the content being transported by the NewsML-G2 Item.

## Urgency

The urgency of the content – in an editorial context – is expressed using an <urgency> rating 1-9, where 1 is the highest urgency and 9 is the least urgent:

```
<contentMeta>
    <urgency>5</urgency>
    …
```

## Dateline information

The dateline associated with content generally consists of a date and the location from where the news event is *reported* . The location is expressed in the <located> element. For photos and videos this is the place where the shutter of the camera was pressed, not necessarily the place visible in the image:

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

Note the place where the subject of the story took place, or the location depicted in an image, is separately expressed using <subject>, which is part of Descriptive Metadata.

## Creator, information source, byline and the credit

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

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

The Information Source for the article may be also given. If no @role is applied, the Information Source provided some information used to create or enhance the content and played no other role. This is the implicit relationship expressed here:

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

NAG uses the <by> element to express the byline:

```
<by>By Meredith Jameson</by>
```

NAG uses the <creditline> element for pictures because the party that must be credited is not necessarily the same as the creator of the content:

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

## Subject Matter

The <subject> element is a repeatable property that expresses what the content is about. This enables the receiver to accurately categorise the content for the end user. In an editorial workflow, the value(s) expressed by <subject> can also ensure that specialist staff see content of interest without having to filter out that which is irrelevant.

In this example, the Subject properties use QCodes from the Controlled Vocabulary of Media Topics NewsCodes that are owned and maintained by the IPTC. The codes are hierarchical with 17 "top-level" terms, (01000000 to 17000000) ranging from Arts to Weather. In this example, the top-level term is 04000000 "economy, business and finance":

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

NAG may indicate the relationship between terms, thus:

```
<subject qcode="medtop:04000000">
    <name>economy, business and finance</name>
</subject>
<subject qcode="medtop:20000350">
    <name>central bank</name>
    <related rel="skos:broader" qcode="medtop:04000000">
</subject>
```

## Slugline

The <slugline> is still used by NAG as a quick human-readable index to a story and its subject matter:

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

## Headline

NAG extracts the headline of the content into the <headline> element as a convenience for receivers:

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

## The Content

The content of a News Item is wrapped in the <contentSet> element. This Quick Start guide focuses on the two most common cases: text and pictures.

## Text

For conveying text NAG uses the element <inlineXML>.

In the example, the IPTC news mark-up language NITF (News Industry Text Format) is used to format the text content. XHTML is also a popular text mark-up choice among NewsML-G2 providers.

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

```
</contentSet>
```

Plain text may be carried using the <inlineData> element using the IANA Media Type of "text/plain" thus:

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

## Pictures

Binary content is conveyed by reference, using the <remoteContent> element.

The <remoteContent> element is repeatable, as more than one rendition of an image may be conveyed in the same News Item, as indicated by the @rendition attribute. NAG sends two renditions of each picture: a large size for high resolution applications and a smaller size for use (for example) on the Web:

```
<remoteContent rendition="rnd:highRes"
   href="./GYI0062134533.jpg" version="1"
   size="346071" contenttype="image/jpeg" width="1500"
   height="1001" colourspace="colsp:AdobeRGB" orientation="1"
   layoutorientation="loutorient:horizontal"/>
<remoteContent rendition="rnd:web"
   href="file:///./GYI0062134533-web.jpg" version="1"
   size="28972" contenttype="image/jpeg" width="480"
   height="320" colourspace="colsp:AdobeRGB" orientation="1"
   layoutorientation="loutorient:horizontal"/>
```

NAG expresses the location of the image as an @href that references a local file path. Other self-explanatory attributes are included as an aid to processing.

For further details on handling pictures and graphics, see the standalone IPTC Quick Start Guide to Conveying Pictures, or the matching chapter in the IPTC NewsML-G2 Guidelines.

### NAG Workflows
*IPTC suggests that details about typical editorial workflows are included, e.g.*

- ❖ Changes of the publication status
- ❖ Embargoes of publication
- ❖ Correction of errors
- ❖ Updates of the content
- ❖ Use of Editorial Notes
- ❖ …

*Such sections should describe how the workflow is exactly expressed by NewsML-G2 properties.*

### NewsML-G2 Implementation Guidelines and Specification

For more comprehensive information about NewsML-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 NewsML-G2 Items in News Messages, and using Planning and Events for news management and fulfilment.

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

# Terms of Use

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 NewsML-G2 Standards and Documentation by visiting www.newsml-g2.org/doc. The version of NewsML-G2 covered by this document is 2.23.

# Contacting NAG about NewsML-G2

NAG contact details here.

## 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