IPTC Standards

NewsML™ G2

(version 2.23)

Quick Start Guide to Conveying Text

Quick Start: Text

Introduction

One of the most fundamental needs of a news organisation is to handle text. This Quick Start Guide covers the basics of a simple NewsML-G2 News Item containing text content.

We recommend reading the Quick Start Guide to NewsML-G2 Basics before this Quick Start Guide to Text.

Example

Below is an example story and supporting information as might be displayed on the journalist's editing screen at a fictional news provider, Acme News and Media (ANM):

Acme News and Media – Content Editing System		
Slugline	US-Finance-Fed	
Created on	2016-11-21 15:21:06	
Source	ANM	
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	
Location / Date	Washington	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. .	

This screen contains nearly all of the information needed to create the NewsML-G2 document below:

Code Listing: NewsML-G2 Text Document

(All Scheme Aliases used in 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 QuickStart 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:acmenews.com:20161121:US-FINANCE-FED"
    version="9"
    standard="NewsML-G2"
    standardversion="2.23"
    xml:lang="en-US">
    <catalogRef
        href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_29.xml" />
    <catalogRef
        href="http://catalog.acmenews.com/news/ANM_G2_CODES_2.xml" />
    <rightsInfo>
        <copyrightHolder uri="http://www.acmenews.com/about.html#copyright">
            <name>Acme News and Media LLC</name>
        </copyrightHolder>
        <copyrightNotice>Copyright 2016 Acme News and Media LLC</copyrightNotice>
    </rightsInfo>
    <itemMeta>
        <itemClass qcode="ninat:text" />
        <provider uri="http://www.acmenews.com/about/" />
        <versionCreated>2016-11-21T16:25:32-05:00</versionCreated>
        <pubStatus qcode="stat:usable" />
```

```xml
      </itemMeta>
      <contentMeta>
         <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.acmenews.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>Staff 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>
```

Document structure

The building blocks of the text 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). The top level (root) element <newsItem> attributes are:

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

This is followed by references to the Catalogs used to resolve QCodes in the Item, and Rights information:

```xml
   <catalogRef
      href="http://www.iptc.org/std/catalog/catalog.IPTC-G2-Standards_29.xml" />
   <catalogRef
      href="http://catalog.acmenews.com/news/ANM_G2_CODES_2.xml" />
   <rightsInfo>
      <copyrightHolder uri="http://www.acmenews.com/about.html#copyright">
         <name>Acme News and Media LLC</name>
      </copyrightHolder>
      <copyrightNotice>Copyright 2016 Acme News and Media LLC</copyrightNotice>
   </rightsInfo>
```

Item Metadata <itemMeta>

Note the three mandatory child elements of the mandatory<itemMeta>:

- ❖ Item Class
- ❖ Provider
- ❖ Version Created

A publication status is also mandatory, but the <pubStatus> element may be omitted, in which case the publication status is "usable". However, it is recommended that the publication status is explicitly given, as in this example. As Acme News & Media is fictional, the Provider property does not use one of the IPTC Provider NewsCodes, and is expressed by a URI:

```xml
<itemMeta>
    <itemClass qcode="ninat:text" />
    <provider uri="http://www.acmenews.com/about.html" />
    <versionCreated>2016-11-21T16:25:32-05:00</versionCreated>
    <pubStatus qcode="stat:usable" />
</itemMeta>
```

Content Metadata <contentMeta>

Administrative Metadata

The administrative properties of the example text story are:

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

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

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

(Note that this is where the story was written, not the place where the subject of the story took place. That would be expressed using <subject>, part of Descriptive Metadata.)

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

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

The Information Source for the article is also given. When used without a @role, <infoSource> is used to denote the person or party that provided the original information on which the content is based. This is the relationship to be expressed here:

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

The default language for the content is given as U.S. English:

```xml
<language tag="en-US" />
```

Descriptive Metadata

In the example, the Subject properties use QCodes from the Controlled Vocabulary of Media Topics NewsCodes that are owned and maintained by the IPTC and expressed as QCodes. Thus:

```xml
<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>
```

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

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

In a similar fashion, the <headline> property will contain the value of the "Headline" field:

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

Complete Content Metadata

```
<contentMeta>
    <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.acmenews.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>
```

Text content choices

Inline XML

The content of the NewsML-G2 document is enclosed by the <contentSet> wrapper. In the example, IPTC's news text mark-up language NITF (News Industry Text Format) is used to format the text content. As an XML standard, it is contained in an <inlineXML> child element of <contentSet>, and uses @contenttype to denote the XML-based standard, using the IANA MIME type.

XHTML is also a popular text mark-up choice among NewsML-G2 providers. As alternatives, the contents of <inlineXML> may be any XML language that can express generic or specialised news information, including SportsML-G2 and rNews. Other languages such as XBRL (Extended Business Reporting Language) may also be used. The content inside <inlineXML> must be valid XML, in other words, it could stand alone as a valid XML document in its own namespace.

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

Inline data

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. The following example uses plain text:

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

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 NewsML-G2 Events for news management and fulfilment.

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

Terms of Use

Copyright © 2016 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 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 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