Quick Start: Pictures and Graphics

Introduction

Image content, including pictures and graphics, can be conveyed in a standard NewsML-G2 document. Picture providers and consumers need a rich vocabulary for descriptive and technical metadata, and for administrative metadata such as rights and usage terms. There is also a long-established use of embedded metadata, such as the IPTC/IIM Fields in JPEG and TIFF files. This Quick Start guide addresses some aspects of embedded metadata, but for a full description of the mapping embedded metadata and IIM fields to G2, this is described in detail in Chapter 21 Mapping Embedded Photo Metadata to NewsML-G2 in the full Guidelines (link provided at the end of this Guide).

The example in this Quick Start guide is a simple but complete document that shows how to implement in NewsML-G2 the frequently-used needs of a professional picture workflow:

- Right and usage instructions
- Descriptive and administrative properties such as Location and Categorisation
- Separate technical renditions of a picture

Read the Quick Guide to G2 Basics before this document. Terms of Use for this document may be read here.

The picture and the metadata used in the example are courtesy of Getty Images. Note that the sample code is NOT intended to be a guide to receiving NewsML-G2 from Getty Images.

About the example

A library picture is provided to customers in three sizes: a large image intended for high resolution and/or large size display, a medium-sized image intended for web use, and a small image for use as a thumbnail or icon. These are three alternative renditions of the same picture and can be contained in a single NewsML-G2 document.

Document structure

The building blocks of the NewsML-G2 document 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 root <newsItem> attributes are:

```
```

Note that this example uses a Tag URI (see TAG URI home page for details)

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

```
<catalogRef href="http://cv.gettyimages.com/nml2catalog4customers-1.xml" />
<rightsInfo>
  <copyrightHolder uri="http://www.gettyimages.com">
    <name>Getty Images North America</name>
  </copyrightHolder>
  <copyrightNotice href="http://www.gettyimages.com/Corporate/LicenseInfo.aspx">(c) 2010 Copyright Getty Images. -- http://www.gettyimages.com/Corporate/LicenseInfo.aspx
    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.</copyrightNotice>
```
Item Metadata <itemMeta>

The <itemClass> property uses a QCode from the IPTC News Item Nature NewsCodes to denote that
the Item conveys a picture.

The Z suffix denotes UTC. Note the <firstCreated> property refers to the creation of the Item, NOT the
content.

Embedded metadata

For many years IPTC metadata fields have been embedded in JPEG or TIFF images files. From 1995 on
the IPTC Information Interchange Model (IIM) defined the semantics of the fields and the technical format
for saving them in image files. In 2003 Adobe introduced a new format for saving metadata, namely XMP
(Extended Metadata Platform), and many IPTC IIM fields were specified as the “IPTC Core” metadata
schema. This defined identical semantics but opened the formats for saving to IIM and XMP in parallel.
Later the “IPTC Extension” metadata schema was added; the defined fields are stored by XMP only. Thus,
many people work with IPTC photo metadata, regardless how they are saved in the files; this is handled by
the software they use.

The transfer of IPTC Photo Metadata fields to NewsML-G2 properties has a focus on the equivalence of
the semantics of fields. The retrieval of the embedded values from the files is a secondary issue and
documents like the Guidelines for Handling Image Metadata, produced by the Metadata Working Group
(http://www.metadataworkinggroup.org/specs/) help in this area.

This Quick Start guide will provide the basics of this mapping, for more details see Chapter 21 Mapping
Embedded Photo Metadata to NewsML-G2 in the main Guidelines, and see the IPTC web site for more
about the IPTC Core and Extension Schemas for XMP
http://www.iptc.org/site/Photo_Metadata/IPTC_Core_&_Extension/

The screen shot on the following page shows the panel for the IPTC Core fields as displayed by Adobe’s
Photoshop CS File Info screen; note the IPTC Extension tab that displays the additional IPTC Extension
metadata.
IPTC Core Metadata fields in the File Info panel of Adobe Photoshop

There are advantages, in a professional workflow, to carrying metadata independently of the binary asset:

- There is no need to retrieve and open the file to read essential information about the picture
- An editor may not have access to the original picture to modify its metadata
- A library picture used to illustrate a news event may have inappropriate embedded metadata.

A situation may arise where the metadata expressed in the NewsML-G2 Item and the embedded metadata in the photo are different. Some providers choose to strip all embedded metadata from objects, to avoid potential confusion. If not, a provider should specify any processing rules in its terms of use.

The IPTC recommends that descriptive metadata properties that exist in the NewsML-G2 Item (in Content Metadata) ALWAYS take precedence over the equivalent embedded metadata (if it exists). These properties include genre, subject, headline, description and creditline.

Content Metadata <contentMeta>

This example shows how embedded metadata from the example picture are translated into NewsML-G2, and includes the equivalent IPTC Core metadata schema property highlighted thus:

IPTC Core Schema equivalent:

Administrative metadata

Timestamp

The <contentCreated> element is used to give the creation date of the picture:

<cont entCr eat ed>2010-10-20T19:45:58-04:00</contentCreated>

Note that this value refers to the creation of the original content; for a scanned picture this is always the date (and optionally the time) of the original photograph. The property type is Truncated Date Time, so that when the precise date-time is unknown, for example for an historic photograph, the value can be truncated (from the right) to a simple date or just a year.

IPTC Core Schema equivalent: Date Created

Creator

The example uses a <creator> element without an identifier, but includes an optional @role that contains a QCode qualifying the creator as a photographer:

<cre at ur e role="crol:photographer"> <nam e>Spencer Platt</name> <rel ed="crol:isA" qcode="gyi bt:sta ff" /> </creator>
The `<related>` child element of `<creator>` further qualifies the photographer as a member of staff (as distinct from, say, a freelance photographer).

**IPCT Core Schema equivalent: Creator**

### Contributor

A `<contributor>` identifies people or organisations who did not originate the content, but have added value to it. In this case, the `@role` hints that the contributor added descriptive metadata:

```xml
<contributor role="ctrol:descrWriter">
  <name>sp/lrc</name>
</contributor>
```

**IPCT Core Schema equivalent: Description Writer**

### Creditline

The `<creditline>` is a natural-language string that must be used by the receiver to indicate the credit(s) for the content, as directed in the business terms agreed with the provider or copyright holder:

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

**IPCT Core Schema equivalent: Credit Line**

### Descriptive metadata

#### Subject

As described in the Quick Start Guide to G2 Basics, the subject matter of content is expressed using the `<subject>` element. The optional `@type` uses the IPTC Concept Nature NewsCodes (recommended scheme alias "cpnat") to indicate the type of concept being expressed. The following example uses a value of "cpnat:event" to indicate that the concept is an Event, and the QCode identifies the Event in the scheme with an alias "gyimeid":

```xml
<subject type="cpnat:event" qcode="gyimeid:104530187" />
```

The provider can use this Event ID to "tag" each of the pictures that relate to this topic, enabling receivers to group them via the Event ID.

The picture of Las Vegas Boulevard illustrates a story about unemployment. This example uses codes and associated `<name>` child elements from the IPTC Media Topic NewsCodes:

```xml
<subject type="cpnat:abstract" qcode="medtop:20000523">
  <name xml:lang="en-GB">labour market</name>
  <name xml:lang="de">Arbeitsmarkt</name>
</subject>

<subject type="cpnat:abstract" qcode="medtop:20000533">
  <name xml:lang="en-GB">unemployment</name>
  <name xml:lang="de">Arbeitslosigkeit</name>
</subject>
```

#### City, State/Province, Country

The `<located>` element in the `<contentMeta>` block describes the place where the picture was created. This may be the same location as the event portrayed in the picture, but this cannot be assumed. The location of the event is logically part of the subject matter – the City, State/Province, Country fields in the IPTC Photo Metadata are defined as "the location shown" – so should use the `<subject>` element. To summarise:

- Use `<located>` to describe where the camera was located when taking the picture.
- Use `<subject>` to describe the location shown in the picture. It is recommended that `@type` is used to indicate the property identifies a geographical area.

The location shown in the example picture is Las Vegas Boulevard. Child elements of `<subject>` may be used to add further details, including:
Quick Start: Pictures and Graphics
NewsML-G2 Implementation Guides
Public Release

- `<name>` gives the place name in plain text, and
- `<broader>` expresses the concept of Las Vegas Boulevard as part of the broader entity of Las Vegas which in turn is part of broader entities of Nevada state and of the United States.

It is recommended that the nature of the concept is indicated by `@type` using a value from the IPTC Concept Nature NewsCodes, in this case that the concept identifies a geographical area:

The completed `<subject>` structure for the geographical information is:

```xml
<subject type="cpnat:geoArea">
    <name>Las Vegas Boulevard</name>
</subject>
<subject type="cpnat:geoArea" qcode="gycon:89109">
    <name>Las Vegas</name>
    <broader qcode="iso3166-2:US-NV">
        <name>Nevada</name>
    </broader>
    <broader qcode="iso3166-1a3:USA">
        <name>United States</name>
    </broader>
</subject>
```

**Keywords**

QCodes and relationship properties are powerful tools, but keywords are still widely used by picture archives. The G2 `<keyword>` property is mapped from the "Keywords" field in XMP. The semantics of "keyword" can vary from provider to provider, but should not present problems in the news industry, which is familiar enough with their use:

```xml
<keyword>business</keyword>
<keyword>economic</keyword>
<keyword>economy</keyword>
<keyword>finance</keyword>
<keyword>poor</keyword>
<keyword>poverty</keyword>
<keyword> gamble</keyword>
```

**IPTC Core Schema equivalent: Keywords**

**Headline, Description**

These two IPTC/IIM fields map directly to elements of the same name in NewsML-G2. Both `<headline>` and `<description>` also have an optional `@role`. The IPTC maintains a set of NewsCodes for Description Role (recommended scheme alias "drol"). In this case, as the description is of a photograph, the role will be "caption". Description is a `Block` type element, meaning it may contain line breaks.

Both elements have optional attributes which may be used to support international use: `@xml:lang`, `@dir` (text direction):

```xml
<headline>Variety Of Recessionary Forces Leave Las Vegas Economy Scarred</headline>
<description role="drol:caption">A general view of part of downtown, including Las Vegas Boulevard, on October 20, 2010 in Las Vegas, Nevada. Nevada once had among the lowest unemployment rates in the United States at 3.8 percent but has since fallen on difficult times. Las Vegas, the gaming capital of America, has been especially hard hit with unemployment currently at 14.7 percent and the highest foreclosure rate in the nation. Among the sparkling hotels and casinos downtown are dozens of dormant construction projects and hotels offering rock bottom rates. As the rest of the country slowly begins to see some economic progress, Las Vegas is still in the midst of the economic downturn. (Photo by Spencer Platt/Getty Images)
</description>
```

**IPTC Core Schema equivalent: Headline**

**Completed `<contentMeta>`**

1 `<broader>` is only available at Power Conformance Level, which is why we set `@conformance` to “power” in `<newsItem>`
A general view of part of downtown, including Las Vegas Boulevard, on October 20, 2010 in Las Vegas, Nevada. Nevada once had among the lowest unemployment rates in the United States at 3.8 percent but has since fallen on difficult times. Las Vegas, the gaming capital of America, has been especially hard hit with unemployment currently at 14.7 percent and the highest foreclosure rate in the nation. Among the sparkling hotels and casinos downtown are dozens of dormant construction projects and hotels offering rock bottom rates. As the rest of the country slowly begins to see some economic progress, Las Vegas is still in the midst of the economic downturn. (Photo by Spencer Platt/Getty Images)
Each `<remoteContent>` wrapper references a separate rendition of the binary resource

Each remote content instance contains attributes that conceptually can be split into three groups:

- **Target resource attributes** enable the receiver to accurately identify the remote resource, its content type and size;
- **Content attributes** enable the processor to distinguish the different business purposes of the content using `@rendition`;
- **Content Characteristics** contain technical metadata such as dimensions, colour values and resolution.

Frequently used attributes from these groups are described below, but note that the NewsML-G2 XML structure that delimits the groups may not be visible in all XML editors. For a detailed description of these attribute groups, see the NewsML-G2 Specification (link provided at the end of this Guide).

**Target Resource Attributes**

This group of attributes expresses administrative metadata, such as identification and versioning, for the referenced content, which could be a file on a mounted file system, a Web resource, or an object within a content management system. G2 flexibly supports alternative methods of identifying and locating the externally-stored content. For this example, the picture renditions are located in the same folder as the NewsML-G2 document.

The two attributes of `<remoteContent>` available to identify and locate the content are Hyperlink (`@href`) and Resource Identifier Reference (`@residref`). Either one MUST be used to identify and locate the target resource. They MAY optionally be used together. Although these two attributes are superficially similar, their intended use is:

- **@href** locates any resource, using an IRI.
- **@residref** identifies a managed resource, using an identifier that may be globally unique.

**Hyperlink (@href)**

An IRI, for example:

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

Or (amongst other possibilities):

```xml
<remoteContent href="file:///./GYI0062134533-web.jpg"/>
```

**Resource Identifier Reference (@residref)**

An XML Schema string, such as:

```xml
<remoteContent residref="/GYI 0062134533-web.jpg"/>
```
Quick Start: Pictures and Graphics

NewsML-G2 Implementation Guides

Public Release

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

Version

An XML Schema positive integer denoting the version of the target resource. In the absence of this attribute, recipients should assume that the target is the latest available version:

Content Type

The MIME type of the target resource:

Size

Indicates the size of the target resource in bytes.

News Content Attributes

This group of attributes of <remoteContent> enables a processor or human-reader to distinguish between different components; in this case the alternative resolutions of the picture. The principal attribute of this group is @rendition, described below.

Rendition

The rendition attribute MUST use a QCode, either proprietary or using the IPTC NewsCode for rendition, which has a Scheme URI of http://cv.iptc.org/newscodes/rendition/ and recommended Scheme Alias of “rnd” and contains (amongst others) the values that we need: highRes, web, thumbnail. Thus using the appropriate NewsCode, the high resolution rendition of the picture may be identified as:

To avoid processing ambiguity, each specific rendition value should be used only once per News Item, except when the same rendition is available from multiple remote locations. In this case, the same value of rendition may be given to several Remote Content elements.

News Content Characteristics

This group of attributes describes the physical properties of the referenced object specific its media type. Text, for example, may use @wordcount). Audio and video are provided with attributes appropriate to streamed media, such as @audiobitrate, @videoframerate. The appropriate attributes for pictures are described below.

Picture Width and Picture Height

The dimension attributes @width and @height are optionally qualified by @dimensionunit, which specifies the units being used. This is a @qcode value and it is recommended that the value is taken from the IPTC Dimension Unit NewsCodes, whose URI is http://cv.iptc.org/newscodes/dimensionunit/ (recommended Scheme Alias is “dimensionunit”)

If @dimensionunit is absent, the default units for each content type are:

<table>
<thead>
<tr>
<th>Content Type</th>
<th>Height Unit (default)</th>
<th>Width Unit (default)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture</td>
<td>pixels</td>
<td>pixels</td>
</tr>
<tr>
<td>Graphic: Still / Animated</td>
<td>points</td>
<td>points</td>
</tr>
<tr>
<td>Video (Analog)</td>
<td>lines</td>
<td>pixels</td>
</tr>
</tbody>
</table>
Quick Start: Pictures and Graphics

NewsML-G2 Implementation Guides

General

The dimensions of the example picture are expressed in pixels, the default, therefore @dimensionunit is not needed:

| width="480" |
| height="2075" |

Picture Orientation

This indicates that the image requires an orientation change before it can be properly viewed, using values of 1 to 8 (inclusive), where 1 (the default) is “upright”: that is the visual top of the picture is at the top, and the visual left side of the picture in on the left.

The application of these orientation values is described in detail in the News Content Characteristics section of the NewsML-G2 Specification document.

The example picture above has an orientation value of 1:

| width="1500" |
| height="1001" |
| orientation="1" |

Layout Orientation

It is possible to calculate the best way to use a picture in a page layout using the combined technical characteristics of Height, Width and Orientation, but many implementers are reluctant to rely on technical characteristics to make editorial judgements (determining whether a video is SD or HD is another example). The @layoutorientation property is a way to express editorial advice on the best way to use a picture in a layout. The value for the example picture is:

| layoutorientation="loutorient:horizontal" |

Values in the Layout Orientation Scheme are:

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>horizontal</td>
<td>The human interpretation of the top of the image is aligned to the long side.</td>
</tr>
<tr>
<td>vertical</td>
<td>The human interpretation of the top of the image is aligned to the short side.</td>
</tr>
<tr>
<td>square</td>
<td>Both sides of the image are about identical, there is no short and long side.</td>
</tr>
<tr>
<td>unaligned</td>
<td>There is no human interpretation of the top of the image.</td>
</tr>
</tbody>
</table>

Picture Colour Space

The colour space of the target resource, and MUST use a QCode. The recommended scheme is the IPTC Colour Space NewsCodes (recommended scheme alias “colsp”) Note the UK English spelling of colour.

| colourspace="colsp:AdobeRGB" |

Colour Depth

The optional @colourdepth property indicates using a non-negative integer the number of bits used to define the colour of each pixel in a still image, graphic or video.

| colourdepth="24" |
Content Hints

At the Power conformance level, the provider is able to express metadata from the target resource\(^2\) as an aid to processing. In this case, the provider has added an `<altId>` – an alternative identifier – for the resource.

Alternative identifiers may be needed by customer systems. The `<altId>` element may optionally be refined using a QCode to describe the context – in this case a "master ID" that is proprietary to the provider. This makes clear the purpose of the alternative identifier. Also note that these Alternative Identifiers are useful only to some other application; they are not intended to be used by THIS NewsML-G2 processor to identify or locate the resource, and note that the provider MUST tell receivers how to interpret alternative identifiers, otherwise their use is meaningless.

```
<altId type="gyiid:masterID">105864332</altId>
```

Note that in this example only the high resolution rendition has an `<altId>`.

Signal

The signal property instructs the G2 processor to process an Item or its content in a specific way. As a child element of itemMeta, the scope of `<signal>` is the whole of the document and/or its contents. If alternative renditions of content have specific processing needs, use `<signal>` as a child element of `<remoteContent>` to specify the processing instructions.

**Completed `<remoteContent>` wrapper**

The `<remoteContent>` wrapping element in full for the “High Res” picture in the example:

```
<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">
    <altId type="gyiid:masterID">105864332</altId>
</remoteContent>
```

**The Complete Picture**

The Listing below is given for completeness. The pictures are identified/located using @href.

```
<newsItem guid="tag:gettyimages.com,2010:GYI0062134533"
    version="3"
    standard="NewsML-G2"
    standardversion="2.15"
    conformance="power"
    xml:lang="en-US"
    xmlns="http://iptc.org/std/nar/2006-10-01"">
    <catalogRef href="http://cv.gettyimages.com/nml2catalog4customers-1.xml" />
    <rightsInfo>
        <copyrightHolder uri="http://www.gettyimages.com">
            <name>Getty Images North America</name>
        </copyrightHolder>
        <copyrightNotice href="http://www.gettyimages.com/Corporate/LicenseInfo.aspx">
            (c) 2010 Copyright Getty Images.
            -
            http://www.gettyimages.com/Corporate/LicenseInfo.aspx
        </copyrightNotice>
        <usageTerms>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:picture">
            2
        </itemClass>
    </itemMeta>
</newsItem>
```

\(^2\) It is not mandatory for the metadata to be extracted from the target resource, but it MUST agree with any existing metadata within the target resource.
Variety Of Recessionary Forces Leave Las Vegas Economy Scarred

A general view of part of downtown, including Las Vegas Boulevard, on October 20, 2010 in Las Vegas, Nevada. Nevada once had among the lowest unemployment rates in the United States at 3.8 percent but has since fallen on difficult times. Las Vegas, has been especially hard hit with unemployment currently at 14.7 percent. Among the sparkling hotels and casinos downtown are dozens of dormant construction projects and hotels offering rock bottom rates. As the rest of the country slowly begins to see some economic progress, Las Vegas is still in the midst of the economic downturn. (Photo by Spencer Platt/Getty Images)
NewsML-G2 Implementation Guidelines and Specification

The full Guidelines for NewsML-G2 Implementers is here:

The NewsML-G2 Specification documents are located here:
http://www.iptc.org/std/NewsML-G2/2.15/specification/

Terms of Use

Copyright © 2014 IPTC, the International Press Telecommunications Council. All Rights Reserved.

This document is published under the Creative Commons Attribution 3.0 license - see the full license agreement at http://creativecommons.org/licenses/by/3.0/. 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.iptc.org and following the link to News Exchange Formats. The version of NewsML-G2 covered by this document is 2.15.

Contacting the IPTC

IPTC, International Press Telecommunications Council
Web address: http://www.iptc.org
Follow us on Twitter: @IPTC
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