CNW Group
XML Format Guide

2009-12-03

Table of Contents

1. Introduction

CNW Group is pleased to introduce an important enhancement of its NewsML based delivery format: Press Releases will now be formatted in XHTML, with true HTML-tables, ready for use in attractive on-line and print applications.

CNW NewsML is a fully-compatible subset of a news-industry standard from the International Press Telecommunications Council (IPTC), providing users with rich, extensible delivery information and metadata for every press release, and the ability to select content based on their local needs.

For example, a CNW NewsML package might include photographs in different resolutions for print or online use. Metadata in CNW NewsML packages includes subject codes identifying items of interest in a press release, such as company stock identifiers and geographical locations, as well as theme and topic codes. At a minimum, each CNW NewsML release produced by CNW Group will include the text of the press release in XHTML format (with news industry in-line markup extensions).

This guide contains the technical specification and examples for CNW Group NewsML XHTML / NITF, together with an extensive index, glossary, list of references, and an overview of controlled vocabularies used for news management.

CNW NewsML XHTML provides users with rich, extensible delivery information and metadata for every news release, and also offers users the ability to select content based on their local needs. For example, a CNW NewsML XHTML package could include photographs in different resolutions for print or online use. Metadata in CNW NewsML XHTML packages includes subject codes identifying items of interest in a news release, such as company stock identifiers and geographical locations, as well as theme and topic codes.

CNW NewsML NITF uses XHTML markup to add structural and semantic information to text to be fully portable, device-independent. Additional information can be included as an option, so that users can generate smart hyperlinks, perform automatic currency conversions, offer intelligent search, and find other ways to add value for their customers.

This guide contains technical specifications and examples for both formats, together with an extensive index, glossary, list of references, and an overview of controlled vocabularies used for news management. For information on delivery protocols, see the *CNW Group Overview of File Formats*, available on the Web at http://xml.newswire.ca

2. CNW NewsML Format

Overview

CNW NewsML is a compatible subset of the NewsML 1.2 specification, designed for simple and efficient news distribution while maintaining full NewsML compatibility and performance. The NewsML format serves three different purposes:

syndication: NewsML encodes information for transmitting news stories from a publisher or distributor to a subscriber, including envelope (routing) information and status.

packaging: NewsML describes how individual news components like stories and pictures fit together into a logical news item, and it provides information for choosing among different versions of the same content (such as the same story in different languages).

metadata: NewsML provides subject codes, rights information, and similar information for lifecycle management, classification, searching, indexing, and other kinds of automated processing.

Content payload such as a news release, may contain multimedia attachments including images (logos and/or photos), video or audio clips. Spreadsheets may be stored in separate files and referenced by links from the NewsML document. Alternatively, CNW NewsML allows plain-text or XML documents, such as CNW NewsML NITF, to be embedded directly inside the NewsML document. Customers have the option of receiving the news release is alternative formats – like html or plain text.

Identifiers

Each CNW NewsML package has a unique identifier, called a NewsML URN (Uniform Resource Name). A NewsML URN looks like this:

```
urn:newsml:newswire.ca:20091202:C1825:1
```

The identifier consists of six parts, separated by colons:

- the string “urn”;
- the string “newsml”;
- a domain name owned by the party creating the identifier (“newswire.ca”);
- a date on which the party owned the domain name, in ISO 8601

Basic format (“20091202”);

- a news item identifier for the content, unique for the combination of the domain name and the date (“C1825” referred to as Cnumber); and
- a positive integer identifying the revision (“1”).

When new versions of a story come out, the revision number will be higher, but the remainder of the identifier will be the same.

Metadata

Metadata—information about news releases—is an important component of CNW NewsML. The information embedded in metadata allows users to process news releases more effectively and classify or locate material of interest. This section introduces the general ideas behind CNW NewsML metadata, while later sections describe the individual types of XML markup in detail.

Metadata Types

CNW NewsML uses seven major types of metadata:

1. transmission data, such as the sender (always CNW) and the date and time of transmission, inside the ***NewsEnvelope*** element;

2. identification data, inside the ***Identification*** element;

3. news management data, such as the date when a news release was first created or when the current revision was created, inside the ***NewsManagement*** element;

4. administrative data, such as the name of the provider (always CNW), inside the ***AdministrativeMetadata*** element;

5. descriptive data, such as subject codes or stock ticker symbols, inside the ***DescriptiveMetadata*** element; and

6. physical data, such as a file size or MIME type, inside the ***ContentItem*** element.

The greatest variety of metadata occurs inside the ***DescriptiveMetadata*** element, which lists industry codes, stock ticker symbols, and other information useful for classification or search and retrieval.

Enumerated Values and Open Values

CNW NewsML uses two different types of metadata values: **enumerated values** and **open values**.

Enumerated values come from a limited range defined in lists maintained by CNW and other parties. They are most useful when there are a relatively small number of values available, the values are all known in advance, and they are not likely to change often. Typical examples of enumerated values include industry codes and release status.

Open values are not defined in advance in a list and new values may appear at any time. They are most useful when there are many values available, when all possible values cannot be known in advance, or when the allowed values are likely to change often. Typical examples of open values include stock symbols or contact information.

In CNW NewsML, enumerated values (also known as a **controlled vocabulary**) appear in many different element types, always attached to an attribute named ***FormalName***, as in the following example:

```
<Status FormalName=″Usable″/>
```

Wherever the FormalName attribute may appear, the optional Scheme and Vocabulary attributes are also allowed, to override any defaults from the Catalog. Vocabulary specifies the URL or URN of the controlled vocabulary, while Scheme provides a subgroup within that vocabulary. For example, to take a status from a custom vocabulary and scheme, you could use markup like this:

```
<Status FormalName="CanUse"
  Vocabulary="http://www.example.org/properties.xml"
  Scheme="status"/>
```

Open values in CNW NewsML are often attached to an element named ***Property***, which looks like this:

```
<Property FormalName="StockSymbol" Value=""TSX:ABC"/>
```

In this element, the ***FormalName*** attribute still contains an enumerated value, but this time the value represents the property's name. The property's value is attached to the ***Value*** attribute, and can take any format at all. As a result, the set of possible property names is listed in advance, in files available through the master catalog, but the set of possible property values can change at any time. In other cases, open metadata is attached to regular XML elements that do not have a ***FormalName*** attribute, as in the following example:

```
<TransmissionId>200912021626CANADANWCANADAPR_C1234</TransmissionId>
```

In these cases, the element name itself functions as the name of the metadata property.

CNW NewsML uses open values for locations, company names, stock ticker symbols, attention lines, contact information, Web sites, and many similar types of metadata.

One special type of enumerated value is the subject code. The ***Subject*** element inside ***SubjectCode*** has an optional ***Scheme*** attribute to show the origin of an enumerated value, as in the following example:

```
<SubjectCode>
  <Subject FormalName="FIN" Scheme="IndustryCode"/>
 </SubjectCode>
```

The following sections describe the components of CNW NewsML. Each section starts with an example showing how that part of the NewsML appears in a typical release. This example release is shown in full in section 4. There then follows a detailed description of all the elements that are permissible in this part of the CNW NewsML. Examples are then given to illustrate elements that were not shown in the initial example.

XML Structure

The top XML element in a CNW NewsML document, ***NewsML***, looks like this:

```
<?xml version="1.0" encoding="UTF-8"?>

<NewsML Version="1.2">
 <Catalog Href="http://xml.newswire.ca/newsml/catalog.xml"/>
 <NewsEnvelope>
  [...]
 </NewsEnvelope>
 <NewsItem>
  [...]
 </NewsItem>
</NewsML>
```

The first line indicates that this file complies with the XML version 1.0 specification and that the character encoding used throughout the file is UTF-8. UTF-8 is an encoding that can represent virtually any type of character and is the default character encoding for XML documents.

There are four required markup components:

- a required ***Version*** attribute that must always be set to the value "1.2";
- a required, non-repeatable ***Catalog*** element pointing to an external, machine-readable catalog of controlled vocabularies used in the document;
- a required, non-repeatable ***NewsEnvelope*** element, containing information about this specific transmission; and
- a required, non-repeatable ***NewsItem*** element, containing metadata and the main content.

The ***Version*** attribute specifies the version of IPTC NewsML to which the document conforms. The ***Catalog*** element is empty and has a required ***Href*** attribute containing a link to an external CNW NewsML catalog, in full (IPTC) NewsML format. The external catalog does not contain the vocabularies in use, but does point to those vocabularies (in separate files) and specify which are the default for each NewsML XML element type. CNW NewsML processing systems are NOT required to retrieve and process this catalog, but it is provided for full IPTC NewsML conformance. The following sections describe the ***NewsEnvelope*** and ***NewsItem*** elements.

News Envelopes

The ***NewsEnvelope*** element holds transmission information, the equivalent of the writing on the outside of a virtual envelope holding the news story, and typically looks like this:

```
<NewsEnvelope>
 <TransmissionId>200409071626CANADANWCANADAPR_C1234</TransmissionId>
 <SentFrom>
  <Party FormalName="CNW"/>
 </SentFrom>
 <DateAndTime>20040907T1626-0400</DateAndTime>
 <Priority FormalName="4"/>
</NewsEnvelope>
```

A ***NewsEnvelope*** can contain four components:

- a required, non-repeatable ***TransmissionId*** element, containing a CNW-specific unique identifier for the transmission;
- a required, non-repeatable ***SentFrom*** element containing a ***Party*** subelement identifying the sender using a formal name (which will always be “CNW”);
- a required, non-repeatable ***DateAndTime*** element, containing the date and time of the transmission in ISO 8601 Basic format;
- a required, non-repeatable ***Priority*** element, specifying the transmission's relative importance from a standard IPTC controlled vocabulary defined at http://www.iptc.org/metadata/mtdta_ts-table01.php.

The ***TransmissionId*** element contains any arbitrary string, as long as it is unique for each transmission. If *exactly* the same information is retransmitted (perhaps because of a communications problem), the transmission ID will be the same, and the optional ***Repeat*** attribute will have a CNW-specific value:

```
<TransmissionId Repeat=”2”>
200912021626CANADANWCANADAPR_C1234</TransmissionId>
```

The ***SentFrom*** element contains a single ***Party*** element:

```
<SentFrom>
 <Party FormalName=”CNW”/>
</SentFrom>
```

Wherever the ***Party*** element appears, it may contain zero or more ***Property*** elements to provide additional information about the person or organization mentioned, as in the following example:

```
<Party FormalName=”ACME News”>
  <Property FormalName=”Phone” Value=”+1 888 555 1234”/>
</Party>
```

User systems can use this element to distinguish CNW NewsML from other types of NewsML received.

News Items

The ***NewsItem*** element is the top of the branch of the CNW NewsML document that holds the actual news content and metadata (which may consist of many different pieces, such as pictures, news, and video in different languages and formats). The XML markup for a news item looks like this:

```
<NewsItem xml:lang="en">
 <Identification>
  [...]
 </Identification>
 <NewsManagement>
  [...]
 </NewsManagement>
 <NewsComponent>
  [...]
 </NewsComponent>
</NewsItem>
```

The ***NewsItem*** contains four markup components:

- an optional ***xml:lang*** attribute specifying the language of the metadata and management information (not the content) using a value in IETF RFC 3066 format;
- a required, non-repeatable ***Identification*** element providing unique identification for this version of the content and metadata;
- a required, non-repeatable ***NewsManagement*** element providing information about the news type, status, and lifecycle; and
- a required, non-repeatable ***NewsComponent*** element containing content, rights information, and descriptive metadata.

The ***NewsComponent*** branch is a major part of CNW NewsML and has its own major section below. The following subsections describe the ***Identification*** and ***NewsManagement*** branches of a news item.

Identification

The ***Identification*** branch of a ***NewsItem*** looks like this:

```
<Identification>
 <NewsIdentifier>
 <ProviderId>newswire.ca</ProviderId>
 <DateId>20091202</DateId>
 <NewsItemId>C1234</NewsItemId>
 <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
 <PublicIdentifier>urn:newsml:newswire.ca:20091202:C1234:1
 </PublicIdentifier>
 </NewsIdentifier>
</Identification>
```

The XML markup for the ***Identification*** element contains three components:

- a required, non-repeatable ***NewsIdentifier*** element containing the

identifier for this version of the news item;

- an optional, non-repeatable ***NameLabel*** element containing a human-oriented nickname for the news item (similar, but not necessarily identical to the slug line); and
- an optional, non-repeatable ***DateLabel*** element containing a presentation-oriented date for the news item (not necessarily in ISO 8601 format).

The optional ***NameLabel*** and ***DateLabel*** contain plain, unstructured text strings in any format chosen by the CNW.

```
<NameLabel>NewsML Sample</NameLabel>
```

```
<DateLabel>Wednesday 2 December 2009</DateLabel>
```

The required ***NewsIdentifier*** element contains five structured components of the unique identifier (all required, non-repeatable sub-elements containing plain text):

- a ***ProviderId*** element containing an Internet domain name owned by CNW on the date specified by the ***DateId*** that follows (the provider id will always be “newswire.ca”);
- a ***DateId*** element containing a date in ISO 8601 Basic format;
- a ***NewsItemId*** element containing a unique name to identify the story (in all versions) – the Cnumber will be used;
- a ***RevisionId*** element containing a positive, non-zero integer distinguishing this version from other versions of the same story (a higher number means a later revision);
- a ***PublicIdentifier*** element containing the previous four components combined into a single NewsML URN; and

The ***RevisionId*** element has two required attributes: ***PreviousRevision***, containing the number of the ***RevisionId*** of the previous version of this news item, or “0” if this is the first version; and ***Update***, which must always have the value “N” for “New”. The content of the ***RevisionId*** attribute should also be incremented when a release is canceled; in this case, the ***Status*** element of ***NewsManagement*** (see below) will have the value “canceled”.

The ***DateId*** has no special meaning for the news release itself: it is simply a date on which CNW owned the domain name used in the ***ProviderId***. Adding a date helps to avoid conflict if a domain name changes ownership. The ***DateId*** can also help with managing ***NewsItemId*** values, since the ***NewsItemId*** needs to be unique only for any given ***ProviderId/DateId*** combination. Note that the date id must be the same for all revisions of a news item: only the revision id may change.

The ***PublicIdentifier*** combines the previous components using the following formula:

```
urn:newsml:ProviderId:DateId:NewsItemId:RevisionId
```

This formula guarantees that each version of a news item will have a unique identifying string.

News Management

The ***NewsManagement*** element inside a news item provides information about the news type, status, and life cycle of a news item. Typical news management information looks like this:

```
<NewsManagement>
 <NewsItemType FormalName="Press Release"/>
 <FirstCreated>20091202T162600-0400</FirstCreated>
 <ThisRevisionCreated>20091202T180000-0400
  </ThisRevisionCreated>
 <Status FormalName="Usable"/>
 <AssociatedWith FormalName=”OtherLanguageEN”
 NewsItem="urn:newsml:newswire.ca:20091202:C4534:1" />
</NewsManagement>
```

The ***NewsManagement*** element contains seven markup components:

- a required, non-repeatable ***NewsItemType*** element providing high-level type for this news item from a standard IPTC controlled vocabulary available at http://www.iptc.org/metadata/mtdta_ts-table01.php: for press releases, the type will be “Press Release”;
- a required, non-repeatable ***FirstCreated*** element containing the date and time that the first version of this news item was created, in ISO 8601 basic format;
- a required, non-repeatable ***ThisRevisionCreated*** element containing the date and time that the current version of this news item was created, in ISO 8601 basic format;
- a required, non-repeatable ***Status*** element providing the status of the news item, such as “Usable” or “Embargoed”, from an IPTC controlled vocabulary available at http://www.iptc.org/metadata/mtdta_ts-table01.php (CNW will normally use the values “Canceled”, “Embargoed”, “Usable”, and “Withheld”);
- an optional, repeatable ***AssociatedWith*** element providing the association of a release to another.
- an optional, repeatable ***StatusWillChange*** element providing notice of scheduled future status changes, such as the end of an embargo, including both the new status (from a controlled vocabulary) and the date and time of the change in ISO 8601 basic format; and
- an optional, non-repeatable ***Urgency*** element providing the relative importance of this news item, from an IPTC standard vocabulary defined at http://www.iptc.org/metadata/mtdta_ts-table01.php, consisting of integers from 1 (highest) to 8 (lowest), together with 9 (user defined); and
- an optional, repeatable ***Property*** element for user-defined news-management metadata.

In the example below, the current status is “Usable,” but it is scheduled to change to “Canceled” at midnight on 31 December 2009.

The ***StatusWillChange*** element contains two required sub-elements: ***FutureStatus***, providing the new status from the same vocabulary as the ***Status*** element above, and ***DateAndTime***, containing the date and time of the future change in ISO 8601 basic format.

```
<StatusWillChange>
 <FutureStatus FormalName=″Canceled″/>
 <DateAndTime>20091231T2359-0400</DateAndTime>
</StatusWillChange>
```

An example of a possible Property element is shown below.

```
<Property FormalName=″AccountCode″ Value=″123456″/>
```

News Components

The ***NewsComponent*** element is a wrapper around a resource like a news release or photograph to hold descriptive information, including its role, the associated newslines, and various kinds of metadata. The NITF news component uses the following basic structure:

```
<NewsComponent Essential=″yes″>
 <Role FormalName=″Main″/>
 <NewsLines>
  [...]
 </NewsLines>
 <AdministrativeMetadata>
  [...]
 </AdministrativeMetadata>
  [...]
 <DescriptiveMetadata>
  [...]
 </DescriptiveMetadata>
 <ContentItem>
  <MimeType FormalName="text/vnd.IPTC.NITF"/>
  <DataContent>
   <nitf>
    [...]
   </nitf>
  </DataContent>
 </ContentItem>
</NewsComponent>
```

The ***Role***, ***NewsLines***, ***AdministrativeMetadata***, and ***DescriptiveMetadata*** elements are all optional.

Sometimes, the same resource will be available in different formats: for example, the body text of a release may be available in both English and French. A ***NewsComponent*** element can itself contain further ***NewsComponent*** elements which can hold more than one version of the same resource. It can also supply information about how to choose between them. See below for an example of a release available in XHTML and IPTC NITF format, each in a separate ContentItem (this is an instructive example only to explain the mechanism that will be used for future expansion; at present CNW Group will deliver only one format in a given NewsML file: either XHTML or NITF):

```
<NewsComponent EquivalentsList=”yes” Essential=”yes”>>
 <BasisForChoice>Language</BasisForChoice>
 <NewsLines>
  [...]
 </NewsLines>
 <AdministrativeMetadata>
  [...]
 </AdministrativeMetadata>
 <DescriptiveMetadata>
  [...]
 </DescriptiveMetadata>

 <NewsComponent xml:lang=”en”>
  <Role FormalName=”Main”/>
  <DescriptiveMetadata>
   <Language FormalName=”en”/>
  </DescriptiveMetadata>
  <ContentItem>
   <MimeType FormalName="text/vnd.IPTC.NITF"/>
   <DataContent>
    [...]
   </DataContent>
  </ContentItem>
 </NewsComponent>

 <NewsComponent xml:lang=”en”>
  <Role FormalName=”Main”/>
  <DescriptiveMetadata>
   <Language FormalName=”en”/>
  </DescriptiveMetadata>
  <ContentItem>
   <MimeType FormalName="application/xhtml+xml"/>
   <DataContent>
    [...]
   </DataContent>
  </ContentItem>
 </NewsComponent>

</NewsComponent>
```

The ***EquivalentsList*** attribute is set to “yes” to indicate that the two content items are alternative versions of the same information. The ***BasisForChoice*** element specifies that the customer's system should choose a version based on the language, either “fr” or “en”.

In other cases, CNW will want to group several related resources together, such as a news release, its formatted contact information and picture or a picture and its caption. When the ***EquivalentsList*** attribute is set to “no”, all of the resources inside the news component are complementary, as in the following example:

```
<NewsComponent EquivalentsList=”no” Essential=”yes”>>
 <NewsLines>
  [...]
 </NewsLines>
 <AdministrativeMetadata>
  [...]
 </AdministrativeMetadata>
 <DescriptiveMetadata>
  [...]
 </DescriptiveMetadata>
```

```
<NewsComponent>
 <Role FormalName=”Main”/>
 <DescriptiveMetadata>
  <Language FormalName=”en”/>
 </DescriptiveMetadata>
 <ContentItem>
  <MimeType FormalName="text/vnd.IPTC.NITF"/>
  <DataContent>
   [...]
  </DataContent>
 </ContentItem>
</NewsComponent>

<NewsComponent EquivalentsList=”no”>
 <Role FormalName=”Supporting”/>
 <DescriptiveMetadata>
  <Language FormalName=”en”/>
 </DescriptiveMetadata>

 <NewsComponent>
  <ContentItem Href=”picture.jpg”>
   <MimeType FormalName=”image/jpeg”/>
  </ContentItem>
 </NewsComponent>

 <NewsComponent>
  <Role FormalNamr=”Caption”/>
  <ContentItem>
   <MimeType FormalName="text/vnd.IPTC.NITF"/>
   <DataContent>
    [...]
   </DataContent>
  </ContentItem>
 </NewsComponent>

</NewsComponent>

<NewsComponent>
 <Role FormalName=”FormattedContactInfo”/>
 <DescriptiveMetadata>
  <Language FormalName=”en”/>
 </DescriptiveMetadata>
 <ContentItem>
  <MimeType FormalName=”text/plain”/>
  <DataContent>
   Test Company
   IT Service Desk
   Phone +1 (416) 123-4567
   email ITSD@testcompany.com
  </DataContent>
 </ContentItem>
</NewsComponent>
</NewsComponent>
```

Over all, then, the ***NewsComponent*** element can contain nine different kinds of markup:

- an ***Essential*** attribute with the value “yes” if this news component is essential to the news item, or “no” (the default) if it is not;

- an ***EquivalentsList*** attribute with the value “yes” if this news component contains alternative versions of the same resource, or “no” (the default) if it contains complementary resources;
- an optional ***xml:lang*** attribute specifying the natural language of the metadata (but not necessarily the content) in IETF RFC 3066 format;
- an optional, non-repeatable ***Role*** element specifying the logical role that this component plays in the news item, from a standard IPTC controlled vocabulary available at http://www.iptc.org/metadata/mtdta_ts-table01.php (CNW will normally use the values “Main”, “Supporting”, and “Caption”);
- an optional, non-repeatable ***BasisForChoice*** element containing an expression specifying how to choose among alternative ***NewsComponent***s or ***ContentItem***s when the parent ***NewsComponent*** has its ***EquivalentsList*** attribute set to “yes”;
- an optional, non-repeatable ***NewsLines*** element containing the headline, byline, slugline, and similar information (see further below);
- an optional, non-repeatable ***DescriptiveMetadata*** element containing language, genre, interest, and subject codes for the news component (see further below); and
- either one or more ***NewsComponent*** elements nested recursively, or one or more ***ContentItem*** elements pointing to the actual news content.

Multiple ***ContentItem*** elements will appear in a ***NewsComponent*** only when they represent exactly the same information in different formats, such as a JPEG and TIFF version of the same photo. Otherwise, each ***ContentItem*** will appear nested in a separate ***NewsComponent*** so that it can have its own news lines and metadata.

The ***BasisForChoice*** element specifies which field a system should use to distinguish alternatives. The three allowed choices for CNW NewsML are as follows:

- “Language” to select ***NewsComponent*** children based on the value of the ***FormalName*** attribute of the ***Language*** element inside ***DescriptiveMetadata***;
- “MimeType” to select ***ContentItem*** children based on the value of the ***FormalName*** attribute of the ***MimeType*** element; and
- “SizeInBytes” to select ***ContentItem*** children based on the content of the ***SizeInBytes*** element inside ***Characteristics.***

The following subsections describe the optional ***NewsLines***, ***AdministrativeMetadata***, and ***DescriptiveMetadata*** elements.

NewsLines

The ***NewsLines*** element contains special text such as headlines, bylines, and copyright lines intended primarily for human readers. News lines are especially important for non-textual content such as photographs, since they are the only source of information for titles and credits. Typical news lines might look like this:

```
<NewsLines>
<HeadLine>Sample with Simple Tables</HeadLine>
 <DateLine>Toronto, Dec. 2</DateLine>
 <CopyrightLine>Copyright CNW Group 2009</CopyrightLine>
 <SlugLine>ON-Test-Release</SlugLine>
 <KeywordLine>green, energy conservation</KeywordLine>
</NewsLines>
```

Up to twelve different kinds of elements can appear:

- an optional, non-repeatable ***Headline*** element, followed by an optional, repeatable ***Subheadline*** element, providing a title and optional subtitles for the news component (this can be a headline for a story or a title for a photo, for example);
- an optional, repeatable ***ByLine*** listing an author of the news component (or similar information like a photo credit);
- an optional, non-repeatable ***DateLine*** element, containing the news component's dateline;
- an optional, repeatable ***CreditLine*** element, containing credit information for the news component (attached to a photo, for example);
- an required, non-epeatable ***CopyrightLine*** element containing a copyright statement for the news component;
- an optional, non-repeatable ***SeriesLine*** element describing a series to which the news component belongs;
- an optional, non-repeatable ***SlugLine*** element providing an abbreviated, human-readable name for the news component;
- an optional, non-repeatable ***KeywordLine*** element containing a whitespace-separated list of unstructured keywords to improve search engine results; and
- an optional, repeatable ***NewsLine*** element containing an extended news line type, such as a caption.

All of these elements except ***NewsLine*** take freeform text rather than controlled vocabularies as their content: with the exception of ***KeywordLine***, they are designed to be presented to humans rather than interpreted by computers. All of these sub-elements can take an ***xml:lang*** attribute describing the natural language of their content using an IETF RFC 3066 format language code, as in the following example:

```
<DateLine xml:lang="fr">TORONTO, le 2 decembre</DateLine>
```

The extended ***NewsLine*** element is slightly different: it specifies its type (from a

controlled vocabulary) using the ***NewsLineType*** subelement, and its content (plain text) using the ***NewsLineText*** element, as in the following example:

```
<NewsLine>
 <NewsLineType FormalName=”Instructions”/>
 <NewsLineText>Not for release before 16:26 EST 2 December
2009</NewsLineText>
</NewsLine>
```

AdministrativeMetadata

The ***AdministrativeMetadata*** element inside a news component contains information on the origin and disposition of a news component. Typical administrative metadata looks like this:

```
<AdministrativeMetadata>
 <FileName>newsml-sample.xml</FileName>
 <Provider>
  <Party FormalName="CNW"/>
 </Provider>
 <Property FormalName="Source" Value="Test Company"/>
</AdministrativeMetadata>
```

The ***AdministrativeMetadata*** element can contain three sub-elements:

- an optional, non-repeatable ***FileName*** element, containing a suggested file name for the news component;
- an optional, non-repeatable ***Provider*** element, specifying the provider of the news component always CNW); and
- an optional, repeatable ***Property*** element containing user-defined administrative metadata.

The ***Provider*** element an empty ***Party*** element that specifies the person or organization concerned from a controlled vocabulary, using the regular ***FormalName*** attribute. The value will always be “CNW” for news releases from CNW.

Property element shown above for Source will appear in most CNW releases. This indicates the original source of the press release.

DescriptiveMetadata

The ***DescriptiveMetadata*** element contains information about the news component itself, particularly information that is useful for classification, indexing, searching, and analysis. Typical descriptive metadata looks like this:

```
<DescriptiveMetadata>
 <Language FormalName=”en”/>
 <Genre FormalName="Press Release"/>
 <SubjectCode>
  <Subject FormalName="04016030"/>
 </SubjectCode>
 <SubjectCode>
  <Subject FormalName="04004003"/>
 </SubjectCode>
 <SubjectCode>
  <Subject FormalName="FIN" Scheme="IndustryCode"/>
 </SubjectCode>
```

```
 <SubjectCode>
  <Subject FormalName="RLT" Scheme="IndustryCode"/>
 </SubjectCode>
 <SubjectCode>
  <Subject FormalName="CPR" Scheme="IndustryCode"/>
 </SubjectCode>
 SubjectCode>
  <Subject FormalName="PDT" Scheme="SubjectCode"/>
 </SubjectCode>
 <NotForDistributionToTheUS/>
 <OfInterestTo FormalName="Business Editors" Scheme="AttentionLine"/>
 <Property FormalName=”CnwNewsMLVersion” Value=”1.1”/>
 <Property FormalName="StockSymbol" Value="TSX:ABC"/>
 <Property FormalName=”CompanyName” Value=”Test Company”/>
 <Property FormalName="GeographyCode" Value="Ontario"/>
 <Property FormalName="WebSite" Value="http://www.testcompany.com"/>
 <Property FormalName="Contact" Value="Joe Smith, CFO, Test Company,
 (416) 123-4567, joe.smith@testcompany.com"/>
 <Property FormalName=”Sedar” Value=”00012345E”/>
 <Property FormalName=”Edgar” Value=”0067890123E”/>
 <OtherInformation FormatName=”someAttribute” Value=”someValue”/>
</DescriptiveMetadata>
```

The ***DescriptiveMetadata*** element can contain the following sub-elements:

- an optional and repeatable ***Language*** element specifying one of the news component's natural (human) languages in ISO 639/IETF 1766 format;
- an optional and repeatable ***Genre*** element specifying a classification code for the news component from the standard IPTC controlled vocabulary available at http://www.iptc.org/metadata/mtdta_ts-table01.php (the normal value will be “Press Release”);
- an optional and repeatable ***SubjectCode*** element containing a ***Subject*** element specifying a formal enumerated code for a subject code in the required ***FormalName*** attribute, together with an identifier for the enumerated codes in the optional ***Vocabulary*** attribute and a subcategory in the optional ***Scheme*** attribute.;
- an optional and non-repeatable ***NotForDistributionToTheUS*** element, specifying a code to alert media that the release is not to be redistributed within the U.S. to their downstream users or subscribers;
- an optional and repeatable ***OfInterestTo*** element, specifying a code for the news component's intended audience, from a CNW enumerated vocabulary;
- an optional, non-repeatable ***DateLineDate*** element containing the pro-forma publication date of the news component in ISO 8601 Basic format; and
- an optional, repeatable ***Property*** element containing CNW descriptive metadata such as locations or stock symbols associated with the news release.
- an optional, non-repeatable ***OtherInformation*** element for internal

CNW use and should be ignored.

The ***Genre*** element will, in this initial release of CNW NewsML, take the value "Press Release" (but may be expanded in later versions of CNW NewsML).

Note that this section contains the bulk of the metadata that can be used to classify the release. Most releases will contain at least some of these elements. See below for details.

Element	Scheme or FormalName	Description
Subject	IndustryCode	Optional, Repeatable Contains a Canada NewsWire's industry code associated with the release.
Subject	SubjectCode	Optional, Repeatable Contains a Canada NewsWire's subject code associated with the release.
Subject	CnwNewsMLVersion	Optional, Non-repeatable Contains a Canada NewsWire's NewsML version code.
Property	StockSymbol	Optional, Repeatable Contains a stock symbol associated with the release.
Property	CompanyName	Optional, Repeatable Contains the name of a company associated with a release.
Property	GeographyCode	Optional, Repeatable Contains a Canada NewsWire's geography code associated with the release.
Property	Website	Optional, Repeatable Contains a website URL associated with the release.
Property	Contact	Optional, Repeatable Contains unformatted contact information associated with a release. Some releases may contain formatted contact information. This will be supplied as a separate NewsComponent.

Property	SEDAR EDGAR	Optional, Repeatable Supports up to 15 EDAR / EDGAR tags.
Property	NoteToEditors	Optional, Repeatable Contains unformatted text representing a non-publishable note for editorial use.
Property	Photo	Optional, Repeatable Contains a web URL for a photograph related to the release but not distributed with it.

ContentItem

The ***ContentItem*** either contains an actual resource or holds a pointer to it. Textual resources, like news stories (including XML files), will usually be included internally; binary resources, like photos or PDF documents, must be included externally. Here is an example of a content item referencing an external file holding an Excel spreadsheet:

```
<ContentItem Href=”year-end.xls”>
 <MimeType FormalName=”application/vnd.ms-excel”/>
 <Characteristics>
  <SizeInBytes>456123</SizeInBytes>
  <Property FormalName=”TargetPlatform” Value=”MSWindows”/>
 </Characteristics>
</ContentItem>
```

Here is an example of a content item including a simple text file internally:

```
<ContentItem>
 <MimeType FormalName=”text/plain”/>
 <DataConent> Test Company
  IT Service Desk
  Phone +1 (416)123-4567
  email ITSD@testcompany.com
 </DataContent>
</ContentItem>
```

CNW currently uses content items for the main news release body (in CNW NITF format), contact information (in plain text), captions (in plain text), pictures (external), and auxiliary files such as spreadsheets (external). If a ***ContentItem*** requires news lines (such as a photo caption) or metadata, it will be wrapped in its own ***NewsComponent***.

The ***ContentItem*** element can contain up to three of the following four markup components:

- an optional ***Href*** attribute containing the relative or absolute URL of the content file;
- a required ***MimeType*** element containing the MIME type of the file;
- an optional ***Characteristics*** element containing a child

element, ***SizeInBytes***, providing the size of the content file and any number of ***Property*** elements; and

- an optional ***DataContent*** element containing the actual textual or XML content.

The ***Href*** attribute and the ***DataContent*** element are mutually exclusive: exactly one of the two must appear, but never both.

CNW NewsML requires the ***MimeType*** element for format identification, since file extensions can be unreliable (and may not be available at all in the case of inline content). The MIME type is especially useful for choosing among different ***ContentItems*** inside a ***NewsComponent*** using the ***BasisForChoice*** element, but it also provides the customer's system with hints about how to process the file.

The optional ***Characteristics*** element contains an optional child element ***SizeInBytes***, providing the size of the content file so that the customer's system can decide whether to download and/or process it, together with an optional and repeatable ***Property*** child element for user-defined characteristics. The file size can be a basis for choosing among different equivalent content items. Here is an example of the ***SizeInBytes*** element:

```
<ContentItem Href=”sample-pic.jpg”>
 <MimeType FormalName=”image/jpeg”/>
 <Characteristics>
  <SizeInBytes>122731</SizeInBytes>
 </Characteristics>
</ContentItem>
```

CNW NewsML XHTML Document Type

A schema for the CNW NewsML XHTML document type is available in:

- W3C XML Schema

This document can be found at http://www.w3.org/1999/xhtml.

3. CNW NewsML NITF/XHTML Format

Overview

CNW Group NewsML format can contain the news item in either NITF or XHTML format. Note: Both the NITF and XHTML format have identical markup for the release body but differ slightly in the header and metadata.

The News Industry Text Format (NITF), from the International Press Telecommunications Council (IPTC), is designed to allow news providers to share loosely-structured textual news components such as news stories or sidebars with rich annotations optionally added. NITF is inspired by the HyperText Markup Language (HTML), and its structure will be familiar to anyone who has used HTML to design Web pages. However, NITF adds (optional) markup of its own for locating and identifying subjects of interest in a news story such as events, locations, people, organizations and currency values.

Extensible Hypertext Markup Language (XHTML) is an extension of the Hypertext Markup Language (HTML).

The NITF and XHTML markup for the DataContent differs only in the header. Additional HTML markup has been added to the NITF format to support tables and stylized text.

With both the NITF and XHTML data content, a news system can perform many kinds of intelligent formatting, searching and classification. Because the documents use XML, customers can convert them to many different formats for print or online delivery.

CNW NITF is CNW's streamlined version of NITF version 3.2. CNW XHTML is XHTML 1.0 Transitional. These formats used together with CNW NewsML, retains all of the expressive power of the original while greatly simplifying the work of rendering, processing, or analyzing news stories.

Controlled Vocabularies

CNW NITF/XHTML allows CNW to annotate news stories with codes from controlled vocabularies. Unlike most existing approaches, where codes are all collected in a single location, CNW's markup optionally allows subject codes to be placed in the main text body at the actual point of reference. The following example annotates a company name with its ticker symbol:

```
<p>The <org value=”CDE”>Test Company</org> announced its
quarterly earnings [...]</p>
```

Because the CNW NITF/XHTML markup shows the exact location where the story mentions the organization with the ticker symbol “CDE”, a search engine can bring the user to the correct point in the story, or a Web site can add a hyperlink to a background page. The attribute name ***value*** is the most common one for specifying

a code from a controlled vocabulary, but the ***location*** element uses the name ***location-code***.

In addition to the code, CNW NITF/XHTML can specify the source of the controlled vocabulary, so that a customer handling stories from several different sources can semantically interpret them. Every CNW NITF/XHTML element that allows a subject code also has an optional attribute for specifying the code's source, most often named ***idsrc***. The format of the source information is not specified: a URL or URN would be the least ambiguous, but might also be verbose:

```
<p>The <org value=”CDE” idsrc=”http://xml.newswire.ca/nitf/company-
codes/”>Test Company</org> announced its quarterly earnings
[...]</p>
```

By default, CNW NITF/XHTML documents will not include source information unless explicitly required by a customer. In that case, a simpler value, like “CNW”, will probably suffice.

The following table lists all of the elements that can be annotated with subject codes, together with the name of the annotation attributes.

Element	***Description***	***Code Attribute***	***Source Attribute***
event	A newsworthy event, like an initial public offering.	value	idsrc
location	A geographical location, real or imaginary.	location-code	code-source
object.title	The title of a human-created object like a book, film, play, or report.	value	idsrc
org	An organization such as a company, government, university, or charity.	value	idsrc
person	A human being, real or imaginary.	value	idsrc

NITF Structure

A NITF document is an XML document that is fully compliant with the XML 1.0 recommendation from the World Wide Web Consortium (W3C).

Many NITF elements—in fact, nearly all of the ones that contain text meant to be read by humans—have an optional ***xml:lang*** attribute that can contain a two-letter ISO 639 language code, optionally followed by a country specifier. The following subsections will not mention this attribute explicitly. For details, see the NITF DTD or schema.

The root element of a NITF document is named ***nitf***, and looks like this:

```
<nitf version="-//IPTC//DTD NITF 3.2//EN"
      change.date="October 10, 2003"
```

```
      change.time="19:30"
      baselang="en">


<head>
  [...]
 </head>
 <body>
  [...]
 </body>
</nitf>
```

The top ***nitf*** element contains up to six markup components:

- a required fixed attribute ***version***, which must have the value "-//IPTC//DTD NITF 3.2//EN";
- a required fixed attribute ***change.date***, specifying the change date of the original IPTC NITF DTD, which must have the value, for example, "October 10, 2003";
- a required fixed attribute ***change.time***, specifying the change time of the original IPTC NITF DTD, which must have the value, for example, "19:30";
- an optional attribute ***baselang***, specifying the base natural language of the news story in ISO 3166 format;
- a required, non-repeatable ***head*** element containing header information; and
- a required, non-repeatable ***body*** element containing the news story itself.

IPTC NITF requires all the first three attributes, and general-purpose NITF processors may fail if they are not present. If a DTD is in use, the values can be defaulted from it.

NITF Header Information

The ***head*** element of a NITF document contains general catalog information about the story, including its title and revision history, and looks like this:

```
<head>
<title>Test Company Announces Fourth Quarter Results and Record
Year End</title>
</head>
```

The ***head*** element contains two markup components:

- a required, non-repeatable ***title*** element specifying the system title (not necessarily the full headline) for the news story; and
- an optional, repeatable ***revision-history*** element providing a record of changes from previous versions of the story.

The ***title*** element contains only plain text. The ***revision-history*** element contains the following components:

- an optional ***name*** attribute specifying the name, initials, or some other means of

identifying the party who made the change;

- an optional ***function*** attribute specifying the role of the party who made the change;
- a required ***norm*** attribute containing the date of the change in ISO 8601 format; and
- a required ***comment*** attribute containing a description of the change.

The value of the function attribute comes from a controlled vocabulary. See below for an example

```
 <revision-history name="dpm" function="writer-author"
  norm="20091202T1626-0400" comment="first draft"/>
 <revision-history name="dpm" function="writer-author"
norm="20091202T1800-0400" comment="added example for head"/>
```

NITF Main Body

The ***body*** element, which holds the main content of the press release, looks like this:

```
<body>
 <body.head>
  [...]
 </body.head>
 <body.content>
  [...]
 </body.content>
 <body.end>
  [...]
 </body.end>
</body>
```

The ***body*** element contains three sub-elements:

- a required, non-repeatable ***body.head*** element holding news lines and rights information;
- a required, non-repeatable ***body.content*** element holding the text of the press release; and
- an optional, non-repeatable ***body.end*** element holding the tagline and bibliography, if present.

The ***body.content*** element holds regular block content such as paragraphs, lists, and tables, as described below (see Block content). The following subsections describe the ***body.head*** and ***body.end*** elements.

body.head

The ***body.head*** element holds the newslines and rights information for the press release, and typically looks like this:

```
<body.head>
 <hedline>
  <hl1>Test Company Software Announces Fourth Quarter Results and
  Record Year End</hl1>
 </hedline>
 <distributor>CNW Group</distributor>
 <dateline>TORONTO, Dec. 2</dateline>
 <abstract>
  [...]
 </abstract>
</body.head>
```

The ***body.head*** element contains up to eight sub-elements:

- a required, non-repeatable ***hedline*** element containing the headline and optional sub-headlines;
- an optional, repeatable ***note*** element holding a publishable or non-publishable note (described under NITF/XHTML Block Content, below);
- an optional, non-repeatable ***rights*** element containing a plain-text statement describing usage and distribution rights;
- an optional, repeatable ***byline*** element providing attribution for the creator or creators of the press release;
- an optional, non-repeatable ***distributor*** element describing the distributor of the press release (i.e., CNW);
- an optional, non-repeatable ***dateline*** element containing the publishable dateline for the press release;
- an optional, non-repeatable ***abstract*** element containing a summary of the press release; and
- an optional, non-repeatable ***series*** element specifying the press release's place in a series of publications.

The ***rights***, ***byline***, ***distributor***, and ***dateline*** elements all contain plain text with no inline markup. The ***hedline*** element contains exactly one ***hl1*** element holding the main headline and zero or more ***hl2*** elements holding sub-headlines: both of these contain regular inline markup (see below). The ***note*** element can also appear as block content, and is described in that section. The ***abstract*** element contains regular block content like paragraphs. Finally, the ***series*** element is empty, with three attributes:

- an optional ***series.name*** attribute containing the name of the series;
- an optional ***series.part*** attribute containing the number of this press release in the series; and
- an optional ***series.totalpart*** attribute containing the total number of press releases in this series.

body.end

The ***body.end*** element contains the tagline and bibliography for the press release, if any, and typically looks like this:

```
<body.end>
 <tagline>Copyright (c) 2009 CNW Group</tagline>
</body.end>
```

The ***body.end*** element can contain two sub-elements:

- an optional, non-repeatable ***tagline*** element containing the tagline for the press release; and
- an optional, non-repeatable ***bibliography*** element containing a list of references for the press release.

The ***tagline*** element contains regular inline content. The ***bibliography*** element can contain only plain text.

XHTML Structure

The root element of an XHTML document is named looks like this:

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:xn="http://www.xmlnews.org/ns/">

<head>
<meta content="text/html" http-equiv="Content-Type"
charset="UTF-8"/>
<title>Test Company Software Announces Fourth Quarter Results
and Record Year End</title>
</head>

<body>

<div class="xn-newslines">
<h1 class="xn-hedline"> Test Company Software Announces Fourth
Quarter Results and Record Year End</h1>
<p class="xn-distributor">Canada NewsWire</p>
<p class="xn-dateline"> TORONTO, Dec. 2</p></div>
<div class="xn-content">
[...]
</div>
</body>
</html>
```

The top ***xhtml*** element contains the following markup components:

- a required, non-repeatable ***html*** element containing an xmlns declaration for the XHTML namespace. The namespace for XHTML is defined to be

http://www.w3.org/1999/xhtml;

- a required, non-repeatable ***head*** element containing header information; and
- a required, non-repeatable ***body*** element containing the news story itself.

XHTML Header Information

The ***head*** element of a NITF document contains general catalog information about the story, including its title and revision history, and looks like this:

```
<head>
 <meta content="text/html" http-equiv="Content-Type" charset="UTF-8"/>
<title>Test Company Announces Fourth Quarter Results and Record
Year End</title>
</head>
```

The ***head*** element contains two markup components:

- a required, non-repeatable ***title*** element specifying the system title (not necessarily the full headline) for the news story; and
- an optional, repeatable ***meta*** element providing information on content type and character set..

The ***title*** element contains only plain text.

Main body

The ***body*** element, which holds the main content of the press release, looks like this:

```
<div class="xn-newslines">
<h1 class="xn-hedline"> Test Company Software Announces Fourth
Quarter Results and Record Year End</h1>
<p class="xn-distributor">Canada NewsWire</p>
<p class="xn-dateline"> TORONTO, Dec. 2</p></div>
<div class="xn-content">
[...]
</div>
</body>
</html>
```

The ***newslines*** element contains three sub-elements:

- a required, non-repeatable ***hedline*** element holding the headline;
- an optional, non-repeatable ***distributor*** element describing the distributor of the press release (i.e. CNW or Canada Newswire); and
- an optional, non-repeatable ***dateline*** element holding the publishable dateline for the press release;

The ***content*** element holds regular block content such as paragraphs, lists, and tables, as described below (see NITF/XHTML Block content).

NITF/XHTML Block content

Block content is information that is normally displayed starting on a new line, like paragraphs, lists, tables, and so on. NITF allows nine elements to appear in a block context in any order, all optional and repeatable. Table elements are added to the NITF format. The following elements are supported in the enhanced NITF/XHTML format:

- the ***p*** element, representing a paragraph;
- the ***hl2*** element, representing a sub-headline anywhere in the story;
- the ***ol*** element, representing an ordered (numbered) list;
- the ***ul*** element, representing an unordered (bulleted) list;
- the ***dl*** element, representing a descriptive (glossary) list;
- the ***bq*** element, representing a block quotation;
- the ***fn*** element, representing a footnote, to be displayed at the bottom of the page or at the end of the story;
- the ***note*** element, representing a note to be displayed at the current location;
- the ***pre*** element, representing preformatted text like an ASCII table or a source-code listing;
- the ***table*** element, representing a table to be displayed at the current location; and

The ***p***, ***ol***, ***ul***, ***dl, pre***, and ***table*** elements are all compatible with HTML version 4 and XHTML.

The ***p*** and ***hl2*** elements contain regular inline content, and the ***pre*** element contains plain text with no sub-elements. The following subsections describe lists, notes, and block quotations, which have special structures.

Lists

NITF supports three list elements:

- ***ul*** for unordered (bulleted) lists;
- ***ol*** for ordered (numbered) lists; and
- ***dl*** for descriptive (glossary) lists.

All three lists come from HTML, but the structure in NewsML is stricter. ***ul*** and ***ol*** both contain one or more ***li*** elements, each of which contains regular block content such as paragraphs. A simple unordered list looks like this:

```
<ul>
 <li>
  <p>First item.</p>
 </li>
 <li>
  <p>Second item.</p>
 </li>
 <li>
  <p>Third item.</p>
 </li>
</ul>
```

Lists can contain other, nested lists inside their items. The ***ol*** element can also contain a ***seqnum*** attribute specifying the starting item number:

```
<ol seqnum=”4”>
 <li>
  <p>Item #4.</p>
 </li>
 <li>
  <p>Item #5.</p>
 </li>
 <li>
  <p>Item #6.</p>
 </li>
</ol>
```

NITF does not specify bullet or numbering styles: those are the responsibility of stylesheets.

The ***dl*** element is different than the other two list elements. The content of a descriptive list consists of a series of pairs of elements, ***dt*** for a term or title, and ***dd*** for the text that follows it (such as a definition or explanation). ***dt*** contains regular inline text, while ***dd*** contains regular block text, as in the following example:

```
<dl>
 <dt>ABC</dt>
 <dd>
  <p>ABC Test Company</p>
 </dd>
 <dt>ECA</dt>
```

```
 <dd>
  <p>EnCana Corporation</p>
 </dd>
 <dt>T</dt>
 <dd>
  <p>TELUS Corporation</p>
 </dd>
</dl>
```

Like any element that contains block content, a single ***dd*** element can contain any number of paragraphs, tables, nested lists, and so on.

Tables

NITF/XHTML supports table elements and the associated sub-elements:

- ***tr*** for defining a row in a table;
- ***th*** for defining a header cell in a table; and
- ***td*** for defining a cell in a table.

The ***th*** and ***td*** elementsboth contain regular block content such as paragraphs. A simple table with three columns and 2 rows looks like this:

```
<table>
 <tr>
   <td>First cell, first row</td>
   <td>Second cell, first row</td>
   <td>Third cell, first row</td>
  </tr>
  <tr>
   <td>First cell, second row</td>
   <td>Second cell, second row</td>
   <td>Third cell, second row</td>
 </tr>
</table>
```

Notes

There are two block elements for notes: ***fn***, representing a footnote, and ***note***, representing a (possibly-publishable) note in the main text flow. Both of these elements contain the same ***body.content*** element that appears at the top of the main story content. A typical footnote looks like this:

```
<fn>
 <body.content>
  <p>SOURCE Test Company</p>
  <p>Web site <virtloc>http://www.testcompany.com</virtloc></p>
 </body.content>
</fn>
```

Any content that can appear inside the main body of a news story can also appear in a note, although in practice a note must never appear inside another note.

The ***note*** element also has two attributes providing more information about the note type:

- a required ***noteclass*** attribute, specifying the purpose of the note; and
- an optional ***type*** attribute, specifying how the note should be handled.

The ***noteclass*** attribute has one of the values “cpyrt” for a copyright note, “editorsnote” for an editorial note, “trademk” for a trademark note, or “undef” for any other kind of note. The ***type*** attribute has the value “std” (the default) for a standards note, “pa” for a publishable advisory, and “npa” for a non-publishable advisory.

Block Quotations

The ***bq*** element, representing a block quotation set off on a new line, normally looks like this:

```
<bq>
 <block>
```

```
  <p>Many forms of Government have been tried, and will be
  tried in this world of sin and woe. No one pretends that
  democracy is perfect or all-wise. Indeed, it has been said
  that democracy is the worst form of Government except all
  those others that have been tried from time to time.</p>
 </block>
 <credit>Winston Churchill</credit>
</bq>
```

The ***bq*** element has two children:

- a required, non-repeatable ***block*** element, containing regular block- level content; and
- an optional, non-repeatable ***credit*** element containing regular inline content.

Note that if a block quotation contains several paragraphs, lists, or tables, they will all appear inside a single ***block*** element.

Inline content

Most locations in NITF/XHTML that allow text also allow inline elements to appear. Most NITF/XHTML inline content is designed to add intelligence to documents, for example, by explicitly flagging the names of people, organizations, and events. Here is a sample of paragraph with only plain text as content:

```
<p>In Toronto, a spokesman for Greg Holung said that he was
recovering well after a June 25 concert in Germany.</p>
```

Here is the same paragraph with inline markup to add extra intelligence:

```
<p>In <location location-code=”l0005611”>Toronto</location>, a
spokesman for <person value=”p3857665”>Greg Holung</person> said
that he was recovering well after a <chron norm=”20090605T000000-
0000”>June 25</chron> concert in
<location location-code=”de”>Germany</location>.</p>
```

The inline markup acts like annotations on words and phrases in the content: a system can ignore them if desired, but it can also use them for intelligent summaries, indexes, searching, and formatting.

The following 16 basic inline elements are allowed:

- ***chron***, for tagging dates and times;
- ***location***, for tagging the names of geographical locations;
- ***money***, for tagging currency amounts;
- ***object.title***, for tagging the titles of books, movies, and other creations;
- ***org***, for tagging the names of companies, governments, and other

organizations;

- ***person***, for tagging the names of people, real or imaginary;
- ***virtloc***, for tagging URLs and e-mail addresses meant to be displayed;
- ***a***, for creating hyperlinks;
- ***em***, for tagging emphasized text;
- ***b***, for tagging bold text;
- ***i***, for tagging italics text;
- ***u***, for tagging underlined text;
- ***br***, for creating a single line break;
- ***hr***, for defining a horizontal line;
- ***lang***, for tagging text in a different language than the surrounding text; and
- ***q***, for tagging direct, inline quotations.

There is no requirement to use all of these inline elements, though some of them (such as ***object.title*** or ***q***) are useful for generating formatting output. A recipient can simply ignore them if desired, but they do add value to documents through the extra intelligence they provide.

chron

The ***chron*** element tags a date or specific time. The contents of the element do not require any specific format, but the optional ***norm*** attribute contains a date and time in ISO 8601 format, with zeros used for any missing information. For example, the following markup tags the meaning of the word “yesterday” as 2 December 2009:

```
<chron norm=”20091202T000000-0000”>Yesterday</chron>, a spokesperson
...
```

location

The ***location*** element annotates a geographical location, real or imagined, mentioned in the body of a press release, and looks like this:

```
<p>Today in <location location-code=”uk.london”>London</location>
[...]</p>
```

The ***location*** element allows only plain text, and uses the following annotation attributes:

- an optional ***location-code*** attribute to provide a subject code for the location; and
- an optional ***code-source*** attribute to specify the controlled vocabulary for the subject code.

money

The ***money*** element contains a currency value and two optional attributes: ***unit***, providing the currency unit in ISO 4217 format, and ***date***, providing the date when the currency amount applied (for conversion purposes):

```
The company will invest <money unit=”USD” date=”200912021T000000-
0000”>$5 million</money> in research and development over the next
two years.
```

object.title

The ***object.title*** element annotates the title of a publication, film, or similar human creation, and looks like this:

```
<p>The success of the film <object.title value=”f58684”>Fahrenheit
911</object.title> has caused the industry to take a second look at
documentaries.</p>
```

The ***object.title*** element allows two annotation attributes:

- an optional ***idsrc*** attribute to specify the controlled vocabulary for the subject code; and
- an optional ***value*** attribute to provide a subject code for the object.

org

The ***org*** element annotates the name of an organization such as a government, company, or university, and looks like this:

```
<p><org value=”t.tsx”>TELUS Corporation</org> announced a partnership
[...]</p>
```

The ***org*** element allows two annotation attributes:

- an optional ***idsrc*** attribute to specify the controlled vocabulary for the subject code; and
- an optional ***value*** attribute to provide a subject code for the organization.

person

The ***person*** element annotates the name of a person, real or imaginary, and looks like this:

```
<p>According to <person value=”p39993”>Bill Gates</person>,
[...]</p>
```

The ***person*** element allows two annotation attributes:

- an optional ***idsrc*** attribute to specify the controlled vocabulary for the subject code; and
- an optional ***value*** attribute to provide a subject code for the person.

virtloc

In NewsML, the ***virtloc*** element contains plain text (no inline markup) representing either an e-mail address or a World Wide Web uniform resource locator (URL):

```
<virtloc>greg@testcompany.com</virtloc>
```

This is an address intended to be shown to the user, rather than simply presented as a link. If the address contains the character "@", it represents an e-mail address; otherwise it represents a URL.

a

The ***a*** element, from HTML, represents a World Wide Web hypertext link embedded in a document. The element has a required ***href*** attribute, containing the Uniform Resource Locator (URL) for the link:

```
<a href="http://xml.newswire.ca/">CNW Group Ltd.</a>
```

em

The ***em*** element, taken from HTML, marks an emphasized phrase (which a stylesheet may render using italics or boldface, for example). The element contains regular inline content:

```
<q>"We will <em>not</em> let our investors down,"</q> said Smith.
```

b

The ***b*** element, taken from HTML, marks a bold phrase. The element contains regular inline content:

```
<q>"We will <b>not</b> let our investors down,"</q> said Smith.
```

i

The ***i*** element, taken from HTML, marks an italics phrase. The element contains regular inline content:

```
<q>"We will <i>not</i> let our investors down,"</q> said Smith.
```

u

The ***u*** element, taken from HTML, marks an underlined phrase. The element contains regular inline content:

```
<q>"We will <u>not</u> let our investors down,"</q> said Smith.
```

br

The ***br*** element, taken from HTML, defines a single line break in the block content. The element is an empty tag:

```
<q>"We will <em>not</em> let our investors down,"</q><br/> said Smith.
```

hr

The ***hr*** element, taken from HTML, defines a horizontal line in the text. The element is an empty tag:

```
<hr/>
```

lang

The ***lang*** element tags words or phrases in a language different than that of the surrounding text. The optional ***value*** attribute specifies the language of the word or phrase:

la beca Walter Meade y otra del <lang value=”en”>Cuban Artist Fund</org>

Stylesheets may use special formatting, like italic text, for foreign words and phrases.

q

The ***q*** element contains an inline quotation without any kind of quotation marks added around it: the stylesheet should handle inserting the correct punctuation for the language, and level of nesting:

```
<q>”Desde el momento en que se anunció la creación de la Latino
Filmmaker Competition, la respuesta fue abrumadora”</q>, dijo
<person>Bernadette Aulestia</person>, vicepresidente de Mercadeo a
Suscriptores y Desarrollo de Marca de HBO.
```

NITF Document Type

A schema for the CNW NewsML NITF document type is available in:

- DTD NITF 3.2

4. Examples

This section contains an example each of a NewsML XHTML news release and a NewsML NITF news release.

Sample of NewsML XHTML news release:

```
<?xml version="1.0" encoding="UTF-8"?>
<NewsML Version="1.2">
 <Catalog Href="http://xml.newswire.ca/newsml/catalog.xml"/>
 <NewsEnvelope>
        <TransmissionId>200911201302CANADANWCANADAPR_C1846</TransmissionId>
        <SentFrom>
                <Party FormalName="CNW"/>
        </SentFrom>
        <DateAndTime>20091120T130200-0500</DateAndTime>
        <Priority FormalName="4"/>
 </NewsEnvelope>
 <NewsItem xml:lang="en">
        <Identification>
                <NewsIdentifier>
                        <WorkOrderNum>6340</WorkOrderNum>
                        <ProviderId>newswire.ca</ProviderId>
                        <DateId>20091120</DateId>
                        <NewsItemId>C1846</NewsItemId>
                        <ClientRefId>401333</ClientRefId>
                        <CompanyRefId>101333</CompanyRefId>
                        <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
                        <PublicIdentifier>urn:newsml:newswire.ca:20091120:C1846:1</PublicIdentifier>
                </NewsIdentifier>
        </Identification>
        <NewsManagement>
                <NewsItemType FormalName="Press Release"/>
                <FirstCreated>20091120T130200-0500</FirstCreated>
                <ThisRevisionCreated>20091120T130200-0500</ThisRevisionCreated>
                <Status FormalName="Usable"/>
        </NewsManagement>
        <NewsComponent Essential="yes">
                <Role FormalName="Main"/>
                <NewsLines>
                        <HeadLine>Sample with simple tables</HeadLine>
                        <DateLine>TORONTO, Nov. 20</DateLine>
                        <SlugLine>BC-Test-Release</SlugLine>
                        <CopyrightLine>Copyright CNW Group 2009</CopyrightLine>
                </NewsLines>
```

```
<AdministrativeMetadata>
        <FileName>200911201302CANADANWCANADAPR_C1846.xml</FileName>
        <Provider>
                <Party FormalName="CNW"/>
        </Provider>
        <Property FormalName="Source" Value="Test Company"/>
        <Property FormalName="DatabaseLine" Value="ME"/>
</AdministrativeMetadata>
<DescriptiveMetadata>
        <Language FormalName="en"/>
        <Genre FormalName="Press Release"/>
        <SubjectCode>
                <Subject FormalName="FIN" Scheme="IndustryCode"/>
        </SubjectCode>
        <SubjectCode>
                <Subject FormalName="ECO" Scheme="SubjectCode"/>
        </SubjectCode>
        <OfInterestTo FormalName="Attention Business Editors"/>
        <Property FormalName="CompanyName" Value="Test Company"/>
        <Property FormalName="GeographyCode" Value="Ontario"/>
        <Property FormalName="Contact" Value="ITSD (Toronto) 1-877-369-4873"/>
        <Property FormalName="CnwNewsMLVersion" Value="1.1"/>
        <Property FormalName="Category" Value="f"/>
</DescriptiveMetadata>
<ContentItem>
        <MimeType FormalName="application/xhtml+xml"/>
        <DataContent>

        <!-- start embedded XHTML document -->

        <html xmlns="http://www.w3.org/1999/xhtml"
            xmlns:xn="http://www.xmlnews.org/ns/">

        <head>
        <meta content="text/html" http-equiv="Content-Type" charset="UTF-8"/>
        <title>Sample with simple tables</title>
        </head>

        <body>

        <div class="xn-newslines">
        <h1 class="xn-hedline">Sample with simple tables</h1>
        <p class="xn-distributor">Canada NewsWire</p>
        <p class="xn-dateline">TORONTO, Nov. 20</p></div>

        <div class="xn-content">
        <p>
            TORONTO, Nov. 20 /CNW/ - The Food Inspection team and Pumpkin Piez
            Inc.are warning the public not to consume certain smoked salmon products
```

```
because they may be contaminated with <i>Listeria monocytogenes</i>.
</p>
<p>
The following Pumpkin Piez brand product is affected by this alert:
</p>
<p>
<br/>
</p>
<table cellpadding="10" cellspacing="0" width="638" border="1">
<tr>
<td width="120" align="center">
<b>Product </b>
</td>
<td width="87" align="center">
<b>Size </b>
</td>
<td width="129" align="center">
<b>Store </b>
</td>
<td width="91" align="center">
<b>Lot #</b>
</td>
<td width="175" align="center">
<b>PACKED ON dates</b>
</td>
</tr>
<tr valign="top">
<td align="left">
Smoked Pumpkin Pie Cheese Log
</td>
<td align="left">
400 g
</td>
<td align="left">
Various stores across Canada
</td>
<td align="left">
C5831DAS
</td>
<td align="left">
OC 16 OC 29 OC 30 NO 03 NO 09 NO 12
</td>
</tr>
</table>
<p>
Salmon Nuggets sold at various weights from the following locations are
also affected by this alert:
</p>
<p>
```

```
</p>
<table cellpadding="10" cellspacing="0" width="637" border="1">
 <tr valign="top">
  <td width="158" align="center">
   <b>Store</b>
  </td>
  <td width="240" align="center">
   <b>Address</b>
  </td>
  <td width="250" align="center">
   <b>Dates offered for sale</b>
  </td>
 </tr>
 <tr valign="top">
  <td align="left">
   Gold Fish Ltd.
  </td>
  <td align="left">
   123 Jones Avenue, Toronto
  </td>
  <td align="left">
   From November 2 to 14, 2009, inclusive
  </td>
 </tr>
 <tr valign="top">
  <td align="left">
   Fish Market West
  </td>
  <td align="left">
   2432 West 16<sup>th</sup> Avenue, Calgary
  </td>
  <td align="left">
   From November 2 to 5, 2009, inclusive
  </td>
 </tr>
 <tr valign="top">
  <td align="left">
   Joe&amp;#8217;s Seafoods
  </td>
  <td align="left">
   77 Park Avenue West, Saskatoon
  </td>
  <td align="left">
   From November 2 to 5, 2009, inclusive
  </td>
 </tr>
 <tr valign="top">
  <td align="left">
```

```
    Nature Foods Market
   </td>
   <td align="left">
    925 Main Street, Winnipeg
   </td>
   <td align="left">
    From November 2 to 8, 2009, inclusive
   </td>
  </tr>
  <tr valign="top">
   <td align="left">
    Good Foods Market Regina
   </td>
   <td align="left">
    5110 8<sup>th</sup> Avenue, Regina
   </td>
   <td align="left">
    From November 2 to 7, 2009, inclusive
   </td>
  </tr>
  <tr valign="top">
   <td align="left">
    Lots of Foods Market Toronto
   </td>
   <td align="left">
    2285 Avenue Road, Toronto
   </td>
   <td align="left">
    From November 2 to 6, 2009, inclusive
   </td>
  </tr>
  <tr valign="top">
   <td align="left">
    Choices Market Halifax
   </td>
   <td align="left">
    5593 Hastings Street, Halifax
   </td>
   <td align="left">
    From November 3 to 7, 2009, inclusive
   </td>
  </tr>
 </table>
 <p>
  There have been no reported illnesses associated with the consumption of
  these products.
 </p>
 <p>
  Food contaminated with <i>Listeria monocytogenes</i> may not look or
```

```
                                    smell spoiled. Consumption of food contaminated with these bacteria may
                                    cause listeriosis, a foodborne illness. Listeriosis can cause high
                                    fever, severe headache, neck stiffness and nausea. Pregnant women, the
                                    elderly and people with weakened immune systems are particularly at
                                    risk. Infected pregnant women may experience only a mild, flu-like
                                    illness, however, infections during pregnancy can lead to premature
                                    delivery, infection of the newborn, or even stillbirth.
                                   </p>
                                   <p>
                                    The manufacturer, Classic Pumpkin Piez Inc. is voluntarily recalling the
                                    affected products from the marketplace. The Food Inspection team is
                                    monitoring the effectiveness of the recall.
                                   </p>
                                   <p>
                                    <b>For further information:</b> ITSD (Toronto) 1-877-369-4873
                                   </p>
                                  </div>
                                  </body>
                                  </html>

                                  <!-- end embedded XHTML document -->

                                  </DataContent>
                          </ContentItem>
                  </NewsComponent>
  </NewsItem>
</NewsML>
```

Sample of NewsML NITF news release:

```
<?xml version="1.0" encoding="UTF-8"?>

<NewsML Version="1.2">
 <Catalog Href="http://xml.newswire.ca/newsml/catalog.xml"/>
 <NewsEnvelope>
  <TransmissionId>200911201302CANADANWCANADAPR_C1846</TransmissionId>
  <SentFrom>
   <Party FormalName="CNW"/>
  </SentFrom>
  <DateAndTime>20091120T130200-0500</DateAndTime>
  <Priority FormalName="4"/>
 </NewsEnvelope>
 <NewsItem xml:lang="en">
  <Identification>
   <NewsIdentifier>
    <WorkOrderNum>6340</WorkOrderNum>
    <ProviderId>newswire.ca</ProviderId>
    <DateId>20091120</DateId>
    <NewsItemId>C1846</NewsItemId>
    <ClientRefId>401333</ClientRefId>
    <CompanyRefId>101333</CompanyRefId>
    <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
    <PublicIdentifier>urn:newsml:newswire.ca:20091120:C1846:1</PublicIdentifier>
   </NewsIdentifier>
  </Identification>

  <NewsManagement>
   <NewsItemType FormalName="Press Release"/>
   <FirstCreated>20091120T130200-0500</FirstCreated>
   <ThisRevisionCreated>20091120T130200-0500</ThisRevisionCreated>
   <Status FormalName="Usable"/>
  </NewsManagement>

  <NewsComponent Essential="yes">
   <Role FormalName="Main"/>
   <NewsLines>
    <HeadLine>Sample with simple tables</HeadLine>
    <DateLine>TORONTO, Nov. 20</DateLine>
    <SlugLine>BC-Test-Release</SlugLine>
    <CopyrightLine>Copyright CNW Group 2009</CopyrightLine>
   </NewsLines>

   <AdministrativeMetadata>
    <FileName>200911201302CANADANWCANADAPR_C1846.xml</FileName>
    <Provider>
     <Party FormalName="CNW"/>
    </Provider>
    <Property FormalName="Source" Value="Test Company"/>
    <Property FormalName="DatabaseLine" Value="ME"/>
   </AdministrativeMetadata>

   <DescriptiveMetadata>
    <Language FormalName="en"/>
    <Genre FormalName="Press Release"/>
```

```
  <SubjectCode>
    <Subject FormalName="FIN" Scheme="IndustryCode"/>
  </SubjectCode>
  <SubjectCode>
    <Subject FormalName="ECO" Scheme="SubjectCode"/>
  </SubjectCode>
  <OfInterestTo FormalName="Attention Business Editors"/>
  <Property FormalName="CompanyName" Value="Test Company"/>
  <Property FormalName="GeographyCode" Value="Ontario"/>
  <Property FormalName="Contact" Value="ITSD (Toronto) 1-877-369-4873"/>
  <Property FormalName="CnwNewsMLVersion" Value="1.1"/>
  <Property FormalName="Category" Value="f"/>
</DescriptiveMetadata>

<ContentItem>
  <MimeType FormalName="text/vnd.IPTC.NITF"/>
  <DataContent>

        <!-- start embedded NITF document -->

        <nitf version="-//IPTC//DTD NITF 3.2//EN"
            change.date="October 10, 2003"
            change.time="19:30"
            baselang="en">

        <head>
        <title>Sample with simple tables</title>
        </head>

        <body>

        <body.head>
        <hedline>
        <hl1>Sample with simple tables</hl1>
        </hedline>
        <distributor>Canada NewsWire</distributor>
        <dateline>TORONTO, Nov. 20</dateline>
        </body.head>

        <body.content>
        <p>
            TORONTO, Nov. 20 /CNW/ - The Food Inspection team and Pumpkin Piez
            Inc.are warning the public not to consume certain smoked salmon products
            because they may be contaminated with <i>Listeria monocytogenes</i>.
          </p>
          <p>
            The following Pumpkin Piez brand product is affected by this alert:
          </p>
          <p>
        <br/>
          </p>
          <table cellpadding="10" cellspacing="0" width="638" border="1">
            <tr>
              <td width="120" align="center">
                <b>Product </b>
              </td>
```

```
    <td width="87" align="center">
      <b>Size </b>
    </td>
    <td width="129" align="center">
      <b>Store </b>
    </td>
    <td width="91" align="center">
      <b>Lot #</b>
    </td>
    <td width="175" align="center">
      <b>PACKED ON dates</b>
    </td>
  </tr>
  <tr valign="top">
    <td align="left">
      Smoked Pumpkin Pie Cheese Log
    </td>
    <td align="left">
      400 g
    </td>
    <td align="left">
      Various stores across Canada
    </td>
    <td align="left">
      C5831DAS
    </td>
    <td align="left">
      OC 16 OC 29 OC 30 NO 03 NO 09 NO 12
    </td>
  </tr>
</table>
<p>
  Salmon Nuggets sold at various weights from the following locations are
  also affected by this alert:
</p>
<p>

</p>
<table cellpadding="10" cellspacing="0" width="637" border="1">
  <tr valign="top">
    <td width="158" align="center">
      <b>Store</b>
    </td>
    <td width="240" align="center">
      <b>Address</b>
    </td>
    <td width="250" align="center">
      <b>Dates offered for sale</b>
    </td>
  </tr>
  <tr valign="top">
    <td align="left">
      Gold Fish Ltd.
    </td>
    <td align="left">
      123 Jones Avenue, Toronto
```

```
  </td>
  <td align="left">
    From November 2 to 14, 2009, inclusive
  </td>
</tr>
<tr valign="top">
  <td align="left">
    Fish Market West
  </td>
  <td align="left">
    2432 West 16<sup>th</sup> Avenue, Calgary
  </td>
  <td align="left">
    From November 2 to 5, 2009, inclusive
  </td>
</tr>
<tr valign="top">
  <td align="left">
    Joe&amp;#8217;s Seafoods
  </td>
  <td align="left">
    77 Park Avenue West, Saskatoon
  </td>
  <td align="left">
    From November 2 to 5, 2009, inclusive
  </td>
</tr>
<tr valign="top">
  <td align="left">
    Nature Foods Market
  </td>
  <td align="left">
    925 Main Street, Winnipeg
  </td>
  <td align="left">
    From November 2 to 8, 2009, inclusive
  </td>
</tr>
<tr valign="top">
  <td align="left">
    Good Foods Market Regina
  </td>
  <td align="left">
    5110 8<sup>th</sup> Avenue, Regina
  </td>
  <td align="left">
    From November 2 to 7, 2009, inclusive
  </td>
</tr>
<tr valign="top">
  <td align="left">
    Lots of Foods Market Toronto
  </td>
  <td align="left">
    2285 Avenue Road, Toronto
  </td>
```

```
            <td align="left">
              From November 2 to 6, 2009, inclusive
            </td>
          </tr>
          <tr valign="top">
            <td align="left">
              Choices Market Halifax
            </td>
            <td align="left">
              5593 Hastings Street, Halifax
            </td>
            <td align="left">
              From November 3 to 7, 2009, inclusive
            </td>
          </tr>
        </table>
        <p>
          There have been no reported illnesses associated with the consumption of
          these products.
        </p>
        <p>
          Food contaminated with <i>Listeria monocytogenes</i> may not look or
          smell spoiled. Consumption of food contaminated with these bacteria may
          cause listeriosis, a foodborne illness. Listeriosis can cause high
          fever, severe headache, neck stiffness and nausea. Pregnant women, the
          elderly and people with weakened immune systems are particularly at
          risk. Infected pregnant women may experience only a mild, flu-like
          illness, however, infections during pregnancy can lead to premature
          delivery, infection of the newborn, or even stillbirth.
        </p>
        <p>
          The manufacturer, Classic Pumpkin Piez Inc. is voluntarily recalling the
          affected products from the marketplace. The Food Inspection team is
          monitoring the effectiveness of the recall.
        </p>
        <p>
          <b>For further information:</b> ITSD (Toronto) 1-877-369-4873
        </p>

      </body.content>
      </body>
      </nitf>

      <!-- end embedded NITF document -->

    </DataContent>
   </ContentItem>
  </NewsComponent>
 </NewsItem>
</NewsML>
```

5. References

[IPTC-Metadata] *Metadata: Subject Reference System & NewsML Topicsets.* International Press Telecommunications Council (IPTC), various dates. http://www.iptc.org/metadata/

[MIME] N. Freed and N. Borenstein. *Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types.* Internet Engineering Task Force (IETF) RFC 2046, November 1996. http://www.ietf.org/rfc/rfc2046.txt

[NITF] *News Industry Text Format (NITF) version 3.2*. International Press Telecommunications Council (IPTC), 14 October 2003. http://www.nitf.org/specifications.php

[NewsML] *NewsML Version 1.2 Functional Specification*. International Press Telecommunications Council (IPTC), 10 October 2003. http://www.newsml.org/IPTC/NewsML/1.2/specification/NewsML_1.2-spec-functionalspec_7.html

[NewsML-URN] A. Coates and D. Allen. *URN Namespace for NewsML Resources.* Internet Engineering Task Force (IETF) RFC 3085, March 2001. http://www.ietf.org/rfc/rfc3085.txt

[URI] T. Berners-Lee at al. *Uniform Resource Identifiers (URI): Generic Syntax.* Internet Engineering Task Force (IETF) RFC 2396, August 1998. http://www.ietf.org/rfc/rfc2396.txt

[UTF-8] F. Yergeau. *UTF-8, a transformation format of ISO 10646.* Internet Engineering Task Force (IETF) RFC 2279, January 1998. http://www.ietf.org/rfc/rfc2279.txt

[XHTML] *XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition): A Reformulation of HTML 4 in XML 1.0.* World Wide Web Consortium (W3C) Recommendation, revised 1 August 2002. http://www.w3.org/TR/xhtml1/

[XML] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and François Yergeau. *Extensible Markup Language (XML) 1.0.* (Third Edition). World Wide Web Consortium (W3C) Recommendation, 4 February 2004. http://www.w3.org/TR/REC-xml/

6. Glossary

a (XML element: NITF)

A hyperlink to an external Web page in NITF inline content.

abstract (XML element: NITF)

A summary of a NITF article.

AdministrativeMetadata (XML element: NewsML)

Information about the origin of a NewsComponent.

ANPA-1312

A modem-based format for newswire distribution, superceded by NITF.

AMPS

An Internet streaming "push" protocol, developed by Acquire Media.

BasisForChoice (XML element: NewsML)

The criterion for choosing among several alternative NewsComponents.

bibliography (XML element: NITF)

A plain-text list of references for an article.

block (XML element: NITF)

A wrapper for the main content of a block quotation (bq).

block content

Paragraph-level XML markup in NITF, such as paragraphs, lists, and tables.

body (XML element: NITF)

The main section of an article.

body.content (XML element: NITF)

The actual content of an article.

body.end (XML element: NITF)

The tagline and bibliography for an article.

body.head (XML element: NITF)

The frontmatter for an article, including news lines.

bq (XML element: NITF)

A block quotation.

ByLine (XML element: NewsML)

A display version of the name of a NewsComponent's author.

byline (XML element: NITF)

A statement of authorship for an article.

Catalog (XML element: NewsML)

A link to an external file listing Metadata vocabularies.

Characteristics (XML element: NewsML)

The physical traits of a ContentItem.

chron (XML element: NITF)

Inline content annotating a date or time.

ContentItem (XML element: NewsML)

A physical news resource like a picture or story.

Contribution (XML element: NewsML)

The role played by the Creator or Contributor to a NewsComponent.

Contributor (XML element: NewsML)

A party who assisted in the creation of a NewsComponent.

controlled vocabulary

A list of unique descriptive codes associated with concepts like subjects, statuses, or parties.

Copyright (XML element: NewsML)

The statement of copyright for a NewsComponent.

CopyrightDate (XML element: NewsML)

The date of copyright for a NewsComponent.

CopyrightHolder (XML element: NewsML)

The party holding copyright for a NewsComponent.

CopyrightLine (XML element: NewsML)

A display version of the copyright statement for a NewsComponent.

Creator (XML element: NewsML)

A party who made a primary contribution to a NewsComponent.

credit (XML element: NITF)

An attribute for a block quotation (bq).

CreditLine (XML element: NewsML)

A display credit for a party who helped to create a NewsComponent.

DataContent (XML element: NewsML)

An inline news resource (textual only).

DateAndTime (XML element: NewsML)

The date and time of transmission for a NewsML package.

DateId (XML element: NewsML)

The date component of a NewsIdentifier, in ISO 8601 Basic format.

DateLabel (XML element: NewsML)

The display version of a NewsItem's publication date.

DateLine (XML element: NewsML)

The display version of a NewsComponents date an/or place of origin.

dateline (XML element: NITF)

The place and/or date an article was created.

DateLineDate (XML element: NewsML)

The date from DateLine in ISO 8601 Basic format.

dd (XML element: NITF)

A description in a glossary list (dl).

DescriptiveMetadata (XML element: NewsML)

Metadata codes for a NewsComponent.

distributor (XML element: NITF)

The party delivering an article.

dl (XML element: NITF)

A glossary list.

dt (XML element: NITF)

A heading in a glossary list (dl).

em (XML element: NITF)

Inline content marking an emphasized phrase.

EndDate (XML element: NewsML)

The finishing date for a specific UsageRight.

FileName (XML element: NewsML)

The suggested file name for a NewsComponent.

FirstCreated (XML element: NewsML)

The date that a NewsItem was first created.

fn (XML element: NITF)

Block content for a footnote to an article.

FutureStatus (XML element: NewsML)

A scheduled release status for a NewsItem.

Genre (XML element: NewsML)

The type classification of a NewsComponent.

Geography (XML element: NewsML)

The area where a usage right is applicable.

head (XML element: NITF)

The official title and revision history of an article.

HeadLine (XML element: NewsML)

A display title for a NewsComponent.

hedline (XML element: NITF)

Container for the headline (hl1) and subheadline (hl2).

hl1 (XML element: NITF)

The primary display title for an article.

hl2 (XML element: NITF)

A secondary title or subheading in an article.

IANA

Internet Assigned Numbers Authority. This association assigns IP addresses and MIME types.

Identification (XML element: NewsML)

Formal identification information for a NewsItem.

inline content

XML markup embedded in the main text flow in NITF, to annotate words or phrases in the text.

IPTC

International Press Telecommunications Council (http://www.iptc.org/).

ISO 4217

An internation standard containing a set of codes for representing world currencies.

ISO 639

An international standard containing a set of codes for representing human languages. NewsML and NITF use the two-letter versions of language names, optionally followed by a country specifier.

ISO 8601

An international standard for representing dates and times. NewsML and NITF use the ISO 8601 Basic format, which appears as year + month + day + "T" + hours + minutes + seconds + ("+" or "-") + hours + minutes, where the last part is the offset from Universal Time Coordinates (Greenwich Mean Time). For example, July 1, 2004 10:00am in New York is "20040701T100000-0400".

KeywordLine (XML element: NewsML)

A list of search keywords for a NewsComponent.

lang (XML element: NITF)

Inline content annotating text in a language different from that of the surrounding text.

Language (XML element: NewsML)

The human language of a NewsComponent's content.

li (XML element: NITF)

An item in an ordered list (ol) or itemized list (ul).

Limitations (XML element: NewsML)

Explicit limitations to the usage rights for a NewsComponent.

location (XML element: NITF)

Inline content annotating a geographical location in an article's text.

metadata

Information about information, such as subject codes for a news story.

MimeType (XML element: NewsML)

The physical data type of a ContentItem (such as "image/tiff").

money (XML element: NITF)

Inline content annotating a monetary value in the text of an article.

NameLabel (XML element: NewsML)

A short, human-readable name for a NewsItem, similar to a slugline.

NewsComponent (XML element: NewsML)

An individual part of a complete NewsItem, such as a story or picture.

NewsEnvelope (XML element: NewsML)

Transmission information for a NewsItem.

NewsIdentifier (XML element: NewsML)

The formal, unique identifier for a NewsItem.

NewsItem (XML element: NewsML)

A complete collection of NewsComponents into a single package.

NewsItemId (XML element: NewsML)

A short title inside a NewsIdentifier (similar to a slugline).

NewsItemType (XML element: NewsML)

A general type classification for a NewsItem.

NewsLine (XML element: NewsML)

Extended news line, such as a caption.

NewsLineText (XML element: NewsML)

Content of an extended news line.

NewsLineType (XML element: NewsML)

Type of an extended news line, from a controlled vocabulary.

NewsLines (XML element: NewsML)

Display lines for a NewsComponent, such as HeadLine and DateLine.

NewsManagement (XML element: NewsML)

Type and status information for a NewsItem.

NewsML

The News Markup Language, an XML format for news syndication, packaging, and metadata from the IPTC (http://www.newsml.org/).

NewsML (XML element: NewsML)

The top level of a NewsML document.

NewsProduct (XML element: NewsML)

The name of a vendor product to which a NewsItem belongs.

NewsService (XML element: NewsML)

The name of a vendor service to which a NewsItem belongs.

NITF

The News Industry Text Format, an XML format for news stories from the IPTC (http://www.nitf.org/).

nitf (XML element: NITF)

The top level of a NITF article.

note (XML element: NITF)

Block content for a note other than a footnote, publishable or non-publishable.

object.title (XML element: NITF)

The title of a human creation, such as a film, book, or report.

OfInterestTo (XML element: NewsML)

The potential audience for a NewsComponent.

ol (XML element: NITF)

An ordered (numbered) list.

org (XML element: NITF)

Inline content annotating an organization, such as a company or government, in article text.

p (XML element: NITF)

Block content for a regular paragraph.

Party (XML element: NewsML)

The name of a person or organization, from a controlled vocabulary.

person (XML element: NITF)

Inline content annotating the name of a person in article text.

plain text

Text in an XML document with no embedded markup.

pre (XML element: NITF)

Block content containing preformatted text, such as an ASCII table.

Priority (XML element: NewsML)

The relative importance of a NewsItem.

Property (XML element: NewsML)

An extended metadata property.

Provider (XML element: NewsML)

The organization making this NewsComponent available.

ProviderId (XML element: NewsML)

The provider field of a NewsIdentifier.

PublicIdentifier (XML element: NewsML)

The NewsIdentifier as a single string.

q (XML element: NITF)

Inline content marking a direct quotation in article text (in place of quotation marks).

revision-history (XML element: NITF)

A single item in the revision history of an article.

RevisionId (XML element: NewsML)

The revision field of a NewsIdentifier.

Role (XML element: NewsML)

The function performed by a NewsComponent within a full NewsItem.

RSS

An XML Web log syndication format

SentFrom (XML element: NewsML)

The sender of a NewsItem, inside the NewsEnvelope.

SentTo (XML element: NewsML)

The recipient of a NewsItem, inside the NewsEnvelope.

series (XML element: NITF)

The name of a series to which an article belongs.

SeriesLine (XML element: NewsML)

A display version of a series to which a NewsComponent belongs.

SizeInBytes (XML element: NewsML)

The size of a ContentItem in bytes.

SlugLine (XML element: NewsML)

A short, unique title for referring to a NewsComponent.

Source (XML element: NewsML)

The origin of a NewsComponent.

StartDate (XML element: NewsML)

The beginning date for a set of UsageRights.

Status (XML element: NewsML)

The release status of a NewsComponent, such as "Usable" or "Embargoed".

StatusWillChange (XML element: NewsML)

A scheduled future change in release status for a NewsComponent.

SubHeadLine (XML element: NewsML)

A secondary display title for a NewsComponent.

Subject (XML element: NewsML)

A code for a subject related to a story.

subject code

A unique identifier for a topic or subject related to a news resource, forming

part of a controlled vocabulary.

SubjectCode (XML element: NewsML)

A code for a subject related to a story (wraps Subject).

tagline (XML element: NITF)

A line containing information such as copyright, appearing after the end of an article.

ThisRevisionCreated (XML element: NewsML)

The date that the current revision of a NewsItem was created.

ticker symbol

A stock market symbol for a publicly-traded company.

title (XML element: NITF)

The official title of an article (not necessarily the headline).

TransmissionId (XML element: NewsML)

A unique identifier for a single transmission of a NewsML document.

ul (XML element: NITF)

An itemized (bulleted) list.

Urgency (XML element: NewsML)

The relative importance of a NewsItem.

URL

Uniform Resource Locator, a unique address for a resource on the World Wide Web (see also URN).

URN

Uniform Resource Name, a unique identifier for a resource, not necessarily on the Internet and not tied to a specific address (seealso URL).

UsageRights (XML element: NewsML)

Rights granted to a specific party for a NewsComponent.

UsageType (XML element: NewsML)

The type of UsageRights granted for a NewsComponent.

virtloc (XML element: NITF)

Inline content marking a virtual location, such as a Web site or e-mail address.

7. Index