



AFP NewsML

Multimedia profile

Evolution 1 (2001-2006)

Revision 5 (April 2006)

CONFIDENTIAL

© 2006 AFP. All rights of reproduction and distribution reserved. You may not copy, reproduce, modify, transmit, publish, display or in any way commercially exploit any of the content of this document without the prior written consent of Agence France-Press.



Table of contents

<u>1.- INTRODUCTION.....</u>	<u>3</u>
1.1.- PURPOSE OF THIS DOCUMENT	3
1.2.- DOCUMENT HISTORY	3
1.3.- RELATED DOCUMENTATION	4
1.4.- SYNTAX.....	4
<u>2.- RUBRIC STRUCTURE</u>	<u>4</u>
2.1.- THE INDEX	5
2.2.- THE MULTIMEDIA DOCUMENT	5
2.2.1.- ENRICHED TEXT	6
2.2.2.- ILLUSTRATION	6
2.3.- PROCESSING MODEL	6
<u>3.- NEWSML GENERIC DESCRIPTION</u>	<u>7</u>
<u>4.- IMPLEMENTATION</u>	<u>8</u>
4.1.- DOCUMENT VALIDATION	8
4.2.- SPECIFIC DETAILS	9
4.2.1.- ABSOLUTE AND RELATIVE URL VALUES	9
4.2.2.- ENCODING	9
4.2.3.- MANDATORY NEWSML ELEMENTS	9
4.2.4.- OPTIONAL ELEMENTS.....	9
4.2.5.- DATES.....	9
4.3.- EVOLUTION OF THIS PROFILE	10
4.4.- THE INDEX	10
4.4.1.- STRUCTURE	10
4.4.2.- EXAMPLE.....	11
4.5.- THE MULTIMEDIA DOCUMENT	12
4.5.1.- GENERIC STRUCTURE.....	12
4.5.2.- ENRICHED TEXT STRUCTURE.....	14
4.5.2.1.- Illustration placeholders.....	15
4.5.2.2.- Hyperlinks.....	16
4.5.2.3.- Organization codes.....	19
4.5.2.4.- Quotes	20
4.5.2.5.- Emphasis	20
4.5.2.6.- Inter-titles.....	20
4.5.2.7.- Lists.....	20
4.5.2.8.- Definition lists.....	20
4.5.2.9.- Tables.....	21
4.5.2.10.- Preformatted text.....	22
4.5.3.- STRUCTURE OF AN ILLUSTRATION	22
4.5.4.- EXAMPLES	23
4.5.4.1.- Simple text document.....	23
4.5.4.2.- Document with a photo.....	24



1.- Introduction

AFP publishes a number of multimedia information products, related to generic news (News On Line / Internet Journal) and major sport events (Olympic Games, etc.).

With such products, each article is presented as a multimedia package, with text, pictures, static graphics, dynamic graphics and video clips. Hyperlinks to other articles from the same product, specific web sites or external documents add depth to the information. The articles may be in French, English, German, Spanish, Portuguese or Arabic, depending of the product.

The multimedia articles are presented in a structured manner: the term "rubric" is used for an ordered collection of articles; depending on the product, the order may be editorially driven or simply reverse-chronological. The number of articles in a rubric is limited. Articles are included or deleted from a rubric in real-time.

Each rubric can be delivered in several formats: HTML (pages are created by the AFP integration team working from client templates), TXT and XML (NewsML). Formated as NewsML, articles are structured without any indication of layout: each client, using XSLT stylesheets or any other XML transform technology, can automatically create a layout adapted to the targeted device.

1.1.- Purpose of this Document

AFP NewsML Multimedia profile of NewsML is a conforming subset of NewsML, tailored to the needs of AFP for delivery of real-time or archive multimedia news.

NewsML is an XML-based standard for all aspects of multimedia news creation, storage and delivery; NewsML has been created by the IPTC (see www.iptc.org and www.newsml.org). AFP actively participates in the evolution of NewsML, and is represented on the IPTC NewsML steering committee. When NewsML is used for multimedia packaging, the text components are marked-up using the NITF standard (again an IPTC standard, see www.nitf.org).

This document introduces recipients of AFP NewsML to the specificities of this new format.

1.2.- Document History

Revision	Release Date	Author	Notes
1	September 6, 2000	L. Le Meur	Internal draft
2	July 17, 2001	L. Le Meur	Public release
3	September 17, 2001	L. Le Meur	Revision 1.1
4	November 13, 2001	L. Le Meur	Revision 1.2
5	January 29, 2002	L. Le Meur	Revision 1.3
6	January 20, 2004	L. Le Meur	Revision 1.4
7	February 17, 2004	L. Le Meur	Revision 1.5 (1.4a) * Addition of the Genre element
8	March 11, 2004	L. Le Meur	Revision 2 * Modification of the video links implementation
9	April 15, 2004	L. Le Meur	Revision 3 * Addition of the SlugLine (NewsLine) * Addition off the CatchLine (NewsLine)

10	March 01, 2005	L. Le Meur	Revision 4 * Implementation of the revision mechanism and addition of an association with the previous version of the NewsItem (4.5.1) * Details on the use of the slugline (4.5.1) * Update of the description of the processing model (2.3) * Change of policy applied to profile updates (4.3)
11	April 6, 2006	L. Le Meur	Revision 5 * Addition of an explicit Byline

1.3.- Related documentation

#	Referenced Document
1	IPTC NewsML Version 1.0 Functional Specification
2	IPTC NewsML Version 1.0 Document Type Definition (DTD)
3	IPTC NITF Version 2.5 Document Type Definition (DTD)
4	IPTC NewsML News Agency Implementation Guideline Version 1.2

1.4.- Syntax

This document describes the elements and attributes as lists. XML elements are described by their names; attributes are associated to their supporting element using the XPath syntax.

Examples:

NewsML @Version	Root element The Version attribute of the NewsML element
Priority/@FormalName	The FormalName attribute in the Priority sub-element of NewsEnvelope.
Location[@HowPresent="Origin"]	The Location element with a HowPresent attribute of value Origin.

2.- Rubric structure

A rubric is represented by an index and a set of multimedia documents.

Practically, the index file and related article files are grouped in a single directory; all images are found in the same directory, or in an 'images' subfolder, depending of the product.

Note that some elements described in this document are implicitly optional. The XML parser software, therefore, must accept

1. the absence of such elements in some NewsML instances,
2. the presence of NewsML elements not described in this document (as long as they are compliant with the NewsML DTD).

Note: this is the standard behavior of XML processing software.



2.1.- The index

The index file lists the articles in the editorial order set by AFP journalists (the *main* news is first in the list), or simply in the reverse-chronological order (the *newer* first), depending of the product.

This file also offers a way to manage the rubric updates (addition, suppression, and update of articles).

The index orders the collections of documents in a rubric by their file names, used as relative URLs.

In order to help the implementation of such a collection, the index adds some information about each document. For simple rubrics, only the document headline is given; but in certain products, some metadata (properties) of the document are extracted and shown in the index. This allows for proper selection and grouping of documents without having to open every document in the collection.

To construct a proper web page, one has to open the associated documents; doing so allows each client to display in such a page any specific information about each document as per their requirements.

The index includes the following parameters:

- Rubric identifier,
- Language,
- Label : most of the rubrics have fixed names ("Top Stories" , "Sports" ...) ; event-related rubrics get variables names (for example "Wimbledon tennis tournament"),
- Creation date,
- Last revision date,
- Language of the child articles,
- Information provider (AFP),
- Ordered list of the documents in the rubric, including:
 - The headline,
 - Some extracted metadata,
 - The file name of the document.

2.2.- The multimedia document

The document includes the following elements:

- Globally unique identifier of the document
- Version number,
- Name (slug),
- Creation date,
- Information provider (AFP),
- Language,
- Genre,
- Rubric identifier,
- Title (headline),
- Sub-title (sub headline, optional),
- Signature (byline),
- Dateline (origin location, optional date, provider),
- Copyright label,
- Slug line,
- Catch line,



- Byline,
- Enriched text,
- One or more illustrations.

(new in revision 4) Revisions may be applied on documents. In such a case the globally unique identifier is stable, and the version number is incremented. There is no relationship between the file names of the different versions of a document.

2.2.1.- Enriched text

The term *enriched text* means a marked-up text that includes:

- Paragraphs,
- Illustration placeholders,
- Preformatted string, sub-titles, tables, lists and other structural elements,
- Organization codes (ISIN & SICOVAM).

“Illustrations” are photos, graphics and (soon) other media objects included in a document. A document can include several such objects; their recommended location is given before a paragraph tag, using a local identifier that leads to another NewsML component identified by the same identifier attribute; a recommended position in the page (left or right) is given for each illustration.

2.2.2.- Illustration

Each illustration object (photo, graphic ...) gets the following parameters:

- Local identifier of the object,
- Provider (credit and optional sources)
- Creator (photographer or graphic artist),
- Title (headline),
- One or more components.

An illustration usually contains several components: the first one - of type text – is the caption; the other have a role of ‘Thumbnail’ (small size), ‘Quicklook’ (medium size), and ‘Preview’ (full screen).

Each component gets the following parameters:

- Role in the illustration (Caption, Thumbnail, QuickLook, Preview),
- Characteristics (format, geometry ...),
- Content URL.

Binary files (jpeg, flash) are usually delivered along with the XML file that represents the document, and found in the same directory or in an ‘images’ subfolder: the URL is then a simple file name, in the current directory. In special cases – especially video clips – the content file is hosted externally; the URL is then an Internet hyperlink.

2.3.- Processing model

The update of the rubric’s files is driven by AFP robots, which deliver in real time any new document file or any new version of the index file, and also suppress any expired file.

The delivery order of files is important for a proper process of the index structure; when a change occurs in a rubric:

- The illustrations associated with new documents are delivered first;
- The text file of the news documents are delivered second;
- Then the new version of the index is delivered;



- At last, the expired files (text and illustrations) are suppressed from the receiving end-point.

The receiving system should be *triggered* by a change of index file, or by the modification of the "last update date" (ThisRevisionCreated) found inside the index document itself. It should then immediately mirror the content of the rubric folder before a new rubric update occurs (the time between updates can be down to 15 seconds), and apply processing on the working folder. No modification, suppression or addition of files should be made on the AFP end-point delivery folder.

An example of processing is then:

1. Open the index file and process it. If you want to use it as an HTML page displaying only the headlines of the documents, an XSLT (or DOM) processing is adequate. If you want to extract information from the associated documents, the included multimedia documents must also be opened and processed accordingly. Each multimedia document is referenced in the index via its file name.
2. Open each multimedia document contained in the index. Process the embedded metadata and textual content. Use the illustration placeholders found in the text to embed in the page the adequate rendition of each illustration. Details about each illustration (metadata and content renditions) are found in specific illustration components. If you want to display an HTML page out of it, an XSLT (or DOM) processing is adequate.

This process can be applied to the NewsML files once, resulting in static HTML pages, or it may be applied each time a user requests a document, via dynamic queries (using jsp, asp or php technologies).

Alternatively the documents may be processed to be stored in a content management system once they have been received. In such a case the expired documents (which have disappeared from the index) may be deleted from the CMS as soon as the index changes. Some customers prefer to keep expired documents online for some time, in order to comments about the news, and delete them automatically after a certain period.

(new in revision 4) When revisions occur on a document, the new version of a document contains a link to the xml file that represents the previous document. The previous versions of the document are not present in the index anymore, but are still available at the receiving end-point, as the AFP robot does not delete them as long as a new version is still usable. This feature is useful for some specific usages of the product (real-time translations, display of the documents in ante-chronological order, i.e. in a *wire mode*).

3.- NewsML generic description

The current version of NewsML is version 1.2 (October 2003). The functional specification document is available on the IPTC site (www.iptc.org), and IPTC manages a specialized Web site, that contains NewsML tutorials and more (www.newsml.org). For added information about NewsML, those sites should be visited; also, an open newsgroup (groups.yahoo.com/group/newsml) allows for Q&A, and communication with the developers of the NewsML standard (including an AFP representative).

The following information is a generic description of NewsML; not all elements are used in the AFP text profile.



The base element of a document is <NewsML>; this container wraps one or several news items. <NewsML> includes a catalog element (<Catalog>) that points to the controlled vocabularies used by AFP, an envelope element (<NewsEnvelope>) that contains the production parameters of the document, and one or several items (<NewsItem>).

<NewsEnvelope> sets the transmission parameters of a document, like a transfer date (<DateAndTime>), origin (<From>), destination (<To>), service (<NewsService>), product (<NewsProduct>), and transmission priority (<Priority>).

<NewsItem> represents a piece of news that handles the document identification parameters (<Identification>), processing (<NewsManagement>), and a main component (<NewsComponent>).

<Identification> contains a set of identification information, like its globally unique identifier (<NewsIdentifier>), and its name (<NameLabel>).

<NewsManagement> sets the document management parameters, like its type (<NewsItemType>), its creation and modification dates (<FirstCreated> and <ThisRevisionCreated>), and its status (<Status>).

<NewsComponent> represents a news component, relative to a <NewsItem>. It handles component metadata (<AdministrativeMetadata>, <DescriptiveMetadata>, <RightsMetadata>), labels (<NewsLines>), and optionally a role (<Role>). Most of all it wraps a list of content items (<ContentItem>), documents or references to documents (<NewsItem> and <NewsItemRef>), or other news components (<NewsComponent>); those different elements can't be mixed.

<AdministrativeMetadata> includes a set of administrative information, like a system identifier, component creators, source and credit of the information.

<DescriptiveMetadata> includes a set of descriptive information, like a genre, subject (using the IPTC subject codes), and its target audience.

<RightsMetadata> includes information about ownership, permissions and usage rights.

<NewsLines> include a set of textual information, like the document title (<HeadLine>), the author signature (<ByLine>), or the original place and date of the news (<DateLine>).

<ContentItem> represents a piece of content, a raw component, a block of enriched text, a jpeg photo, png graphic, or a video clip. The element has physical attributes, like a size, format, content item geometry etc. Data is included or pointed to via an URL.

4.- Implementation

The AFP NewsML Multimedia profile current implementation is based on NewsML1.0 and NITF2.5.

Those were the current standard versions when this profile was originally designed in year 2000, and AFP had since no need to add any feature not supported by those versions.

4.1.- Document validation

The XML documents distributed to AFP customers have no type declaration (DOCTYPE); AFP makes its best effort to distribute valid documents.



If you choose to validate incoming documents, you can copy locally the NewsML and NITF DTDs from the IPTC web site, and create "on the fly" a type declaration in the xml instances. The proper format is (/dtd is an example of local URL):

```
<!DOCTYPE NewsML SYSTEM "/dtd/NewsMLv1.0.dtd" [  
<!ENTITY % nitf SYSTEM "/dtd/nitf-2-5.dtd">  
%nitf;  
>
```

4.2.- Specific details

4.2.1.- Absolute and relative URL values

The links between AFP multimedia documents are relative links, valid if the client publication structure (directory tree) reflects the delivery structure. If this is not the case, it's up to the client to modify the URLs in order to reflect the publication structure. AFP manages the synchronization of the deletion of links before pointed documents are suppressed in the delivery structure. AFP can't guarantee the validity of links if the publication content doesn't mirror the delivered content.

Resource files representing dynamic graphics and video clips are not deleted from the delivery space by AFP. Those files may be deleted by the client system three weeks after their publication, as not links will point any more at these files from AFP multimedia documents after this period.

Some links point to external resources, such as Web pages: the lifetime of those links can't be guaranteed by AFP.

4.2.2.- Encoding

AFP currently supports **UTF-8** (sports products in all languages), **iso-8859-1** (European languages) and **windows-1256** (Arabic) encodings for its multimedia products.

At some point in the future, **UTF-8** will be the unique globally supported encoding.

4.2.3.- Mandatory NewsML elements

Some elements that are present in the document don't seem of great interest at first glance, and always contain the same value. But these elements are mandatory in the NewsML DTD, and so must be assigned a value.

4.2.4.- Optional elements

Note that in certain occasions some optional elements are not present in the document instances. Therefore the XML processing software must accept the absence of such elements and use the default value described in this documentation.

4.2.5.- Dates

IPTC uses the **ISO 8601:2000** basic format as date format; dates may be UTC based or represent a local time.

Example: 20011205T150000Z stands for year 2001, fifth of November, three pm UTC time (using the Z [Zulu] designator as UTC flag); this is equivalent to 20011205T160000+0100 or 20011205T130000-0200.



Note: For more information, check: <http://www.pvv.org/~nsaa/8601v2000.pdf>

4.3.- Evolution of this profile

AFP is committed to offer to its customers the best in news content and the most powerful features available in the information community. The side effect of this is that AFP profiles have to evolve when new features are integrated. To allow for a proper management of evolutions on the client side, user support teams will be alerted each evolution up to three months before it effectively takes place. In order to be alerted properly, the user support team should make sure that their mail address is known from the AFP client-support team (mailto:client-support@afp.com).

Evolutions will be as much as possible backward compatible, so that the recipient processing software treats new instances as usual, and can disregard the new elements as long as they are compliant with the NewsML DTD (this behavior is standard for XML parsers). In such a case the change notice will be sent *one month* before the update takes place. AFP does not plan to achieve such an evolution more than twice a year.

A major revision will occur when AFP decides to adopt the future NewsML v2 model (not before the end of 2005). In case of such a major revision of the profile, the change notice will be sent *three months* before the update, and customers will have the possibility to keep the previous profile for one year. After this period they will be invited to move to the new profile, as the old one will be deprecated.

In the following description, some paragraphs get an "Evolution" note, where future modifications are described; doing so, we hope that implementers will be able to upgrade easily their software as soon as they are activated.

4.4.- The index

4.4.1.- Structure

The index includes the following parameters:

1. NewsML/Catalog/@Href: AFP controlled vocabularies catalog ; this catalog contains the values and description of all the elements in which a limited set of names is allowed,
2. NewsML/NewsEnvelope/DateAndTime: delivery date,
3. NewsML/NewsItem: item that represents the rubric; see below.

The main **NewsItem** element describes the rubric:

4. NewsItem/Identification/NewsIdentifier/ProviderId: provider ('afp.com'),
5. NewsItem/Identification/NewsIdentifier/DateId: document creation date,
6. NewsItem/Identification/NewsIdentifier/NewsItemId: identifier,
7. NewsItem/Identification/NewsIdentifier/RevisionId: revision ('1'),
8. NewsItem/Identification/NewsIdentifier/RevisionId/@PreviousRevision: previous revision ('0'),
9. NewsItem/Identification/NewsIdentifier/RevisionId/@Update: update flag ('N'),
10. NewsItem/Identification/NewsIdentifier/PublicIdentifier: global identifier of the document (NewsML URN),
11. NewsItem/Identification/NameLabel: name of the rubric,
12. NewsItem/NewsManagement/NewsItemType/@FormalName: document type ('News'),
13. NewsItem/NewsManagement/FirstCreated: creation date,



14. NewsItem/NewsManagement/ThisRevisionCreated: last update date (equal creation date),
15. NewsItem/NewsManagement/Status/@FormalName: status ('Usable'),
16. NewsItem/NewsComponent: main component; see below.

The main **NewsComponent** element describes the documents that belong to the rubric:

17. NewsComponent/AdministrativeMetadata/Provider/Party/@FormalName: provider of the documents ('AFP'),
18. NewsComponent/DescriptiveMetadata/Language/@FormalName: language used in the documents, ISO 639 format (2 letters),
19. NewsComponent/NewsComponent: components; see below.

Each included **NewsComponent** briefly describes one item in the rubric:

20. NewsComponent /NewsLines/HeadLine: headline,
21. NewsComponent /DescriptiveMetadata/Property/@FormalName: name of a property,
22. NewsComponent /DescriptiveMetadata/Property/@Value: value of a property,
23. NewsComponent /NewsItemRef/@NewsItem: local URL of the file that contains the document (*.xml),

4.4.2.- Example

```
<?xml version="1.0" encoding="iso-8859-1"?>
<NewsML>
  <Catalog Href="http://www.afp.com/dtd/AFPCatalog.xml"/>
  <NewsEnvelope>
    <DateAndTime>20010608T093354Z</DateAndTime>
  </NewsEnvelope>
  <NewsItem>
    <Identification>
      <NewsIdentifier>
        <ProviderId>afp.com</ProviderId>
        <DateId>20040123</DateId>
        <NewsItemId>mmd--english--webwires--europe--uk</NewsItemId>
        <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
        <PublicIdentifier>urn:newsml:afp.com:20040123:mmd--english--webwires--europe--
uk:1</PublicIdentifier>
      </NewsIdentifier>
      <NameLabel>UK news</NameLabel>
    </Identification>
    <NewsManagement>
      <NewsItemType FormalName="News"/>
      <FirstCreated>20040123T173354Z</FirstCreated>
      <ThisRevisionCreated>20040123T173354Z </ThisRevisionCreated>
      <Status FormalName="Usable"/>
    </NewsManagement>
    <NewsComponent>
      <AdministrativeMetadata>
        <Provider>
          <Party FormalName="AFP"/>
        </Provider>
      </AdministrativeMetadata>
      <DescriptiveMetadata>
        <Language FormalName="fr"/>
        <OfInterestTo FormalName="mmd--english--webwires--europe--uk"/>
      </DescriptiveMetadata>
      <NewsComponent>
        <NewsLines>
          <HeadLine>Lib Dem sparks row over Palestinian suicide bomber claim</HeadLine>
        </NewsLines>
        <NewsItemRef NewsItem="010608090901.lchbih5s.xml"/>
      </NewsComponent>
    </NewsComponent>
  </NewsItem>
</NewsML>
```



```

</NewsComponent>
<NewsComponent>
  <NewsLines>
    <HeadLine>Thai poultry meat banned over bird flu fears</HeadLine>
  </NewsLines>
  <NewsItemRef NewsItem="010608064211.kgsqm8jk.xml"/>
</NewsComponent>
.....
<NewsComponent>
  <NewsLines>
    <HeadLine>Blair will survive Kelly affair, but Iraq row will go on: Cook</HeadLine>
  </NewsLines>
  <NewsItemRef NewsItem="010608073758.y4ome30f.xml"/>
</NewsComponent>
</NewsComponent>
</NewsItem>
</NewsML>

```

4.5.- The multimedia document

A multimedia document is a <NewsML> structure that contains a <NewsItem>; its location is found in the index of its rubric (NewsItemRef); the <Identification>, <AdministrativeMetadata>, <DescriptiveMetadata> headers contain the document metadata, <NewsLines> bring publishable textual elements. The NewsItem contains one or more <NewsComponent> elements that represent the text components and associated illustrations; the recommended placement of those objects around text is set in the enriched text component.

4.5.1.- Generic structure

The document includes the following parameters:

1. NewsML/Catalog/@Href: AFP controlled vocabularies catalog,
2. NewsML/NewsEnvelope/DateAndTime: delivery date,
3. NewsML/NewsItem: item that represents the document; see below.

The main **NewsItem** element describes the document:

4. NewsItem/Identification/NewsIdentifier/ProviderId: provider ('afp.com'),
5. NewsItem/Identification/NewsIdentifier/DateId: creation date,
6. NewsItem/Identification/NewsIdentifier/NewsItemId: identifier,
7. NewsItem/Identification/NewsIdentifier/RevisionId: version number ('1' for the first version),
8. NewsItem/Identification/NewsIdentifier/RevisionId/@PreviousRevision: previous version number ('0' for the first version),
9. NewsItem/Identification/NewsIdentifier/RevisionId/@Update: update flag (always 'N'),
10. NewsItem/Identification/NewsIdentifier/PublicIdentifier: global identifier of the document (NewsML URN),
11. NewsItem/Identification/NameLabel: name of the document ("slug"),
12. NewsItem/NewsManagement/NewsItemType/@FormalName: document type (always 'News'),
13. NewsItem/NewsManagement/FirstCreated: date of creation of the first version of the document,
14. NewsItem/NewsManagement/ThisRevisionCreated: date of creation of this version of the document (**modified in revision 4**),
15. NewsItem/NewsManagement/Status/@FormalName: status (always 'Usable'),



16. NewsItem/NewsManagement/DerivedFrom/@NewsItem: **Optional**, appears only if the case of a version number higher than '1'. File name of the previous version of the document (*new in revision 4*).
17. NewsItem/NewsComponent: main component; see below.

The main NewsComponent element describes the document:

18. NewsComponent/NewsLines: textual information, directly publishable; see below,
19. NewsComponent/AdministrativeMetadata: administrative metadata; see below,
20. NewsComponent/DescriptiveMetadata: descriptive metadata; see below,
21. NewsComponent/NewsComponent: component (enriched text or illustration), **repeatable**; see below.

NewsLines describe some directly publishable textual information:

22. NewsLines/HeadLine: Headline (title) of a story.
23. NewsLines/SubHeadLine: **Optional**. May represent some advisory.
24. NewsLines/ByLine: **Optional**. Byline (name of the author) of a story (*new in revision 5*).
25. NewsLines/DateLine: Dateline of a story; a location and optionally a date of origin of the news, plus the name of the news provider.
26. NewsLines/CopyrightLine: Copyright information.
27. NewsLines/SlugLine: Slugline, as originally used in news feeds (a set of hyphen separated keywords, plus an optional version field).

Optionally (in some languages only, e.g. German), a specific NewsLine appears:

28. NewsLines/NewsLine/NewsLineType[@FormalName="CatchLine"]: identifies a catch-phrase label.
29. NewsLines/NewsLine/NewsLineText: text of the catch-phrase. This is currently a copy of the first sentence of the story's first paragraph.

Note about the Slug and SlugLine:

Slugs are designed to differentiate dispatches, organise the wires and facilitate computer searches. Slugs may not exceed 64 characters, including hyphens. No story may carry the same slug as another story within the 24-hour cycle, unless the second is marked with the attribute new-series.

A slug can have up to four elements. The first is the master slug. It may be used alone. Almost all country names are master slugs (occasionally, a shorter word is used to save space, such as Dutch for Netherlands). Other master slugs include continents and geographical regions (such as Africa, Caribbean, Mideast), international bodies (FAO, NATO, WBank), important themes (Aviation, Disarm, Warcrimes) and economic terms (Forex, Stocks).

The second element, separated from the master slug by a hyphen, is the theme slug or common key word. There are about 30 commonly used theme slugs. Again, the list is not exhaustive. Themes include agriculture, animals, arts, cinema, crime, education, energy, environment, fashion, health, justice, labour, markets, media, music, people, politics, population, quake, refugees, religion, rights, science, social, space, terrorism, tourism, transport, unrest, vote.

The third element is a story identifier or keyword freely chosen. On rare occasions a slug may go to a fourth word, provided it falls within 64 characters.

Sports stories have a different syntax, since they need to be classified by sports, tournament, league etc., as well as by country. Results also need to be distinguished from text. The slug starts with the sport. The second element is a three-letter country code, or the region, or the specific competition (World, Cup etc). The third element is the story identifier. This can go to a fourth word if necessary.



Slug elements are followed by attributes. There are two kinds of attribute. One refers to the story content, the other to the story's hierarchical place in the file. Content attributes always come first in the version field. The exhaustive list is: **advancer, analysis, background, chrono, facts, feature, obit, profile, points, quotes, reax, res, score, scores, standings, table, text**. The hierarchy attributes are: **lead, new-series, roundup**. They always come second, separated from the content attributes by a hyphen. If the story has been scheduled on an advisory, the attribute `sched` comes before the content and the hierarchy attributes. Example: Britain-politics-Blair, sched-profile-lead

The **AdministrativeMetadata** element defines the administrative metadata:

30. AdministrativeMetadata/Provider/Party/@FormalName: information provider (always 'AFP').

Evolution: the name and position of the creator of the document may be added in the future as `AdministrativeMetadata/Creator/Party/@FormalName`.

The **DescriptiveMetadata** element defines the descriptive metadata:

31. DescriptiveMetadata/Language/@FormalName: language used in the rubric, ISO 639 format (2 letters).
32. DescriptiveMetadata/OfInterestTo/@FormalName: name of the rubric the document belongs to; this allows for some categorisation of the documents without using the index.

Evolution:

- The editorial Genre (Current (default), Prev, Lead, Analysis, etc.) may be added in the future as `DescriptiveMetadata/Genre/@FormalName`. It has been withheld since revision 3 of this document, as the AFP used values outside of the IPTC controlled vocabulary.
- The editorial version of the news story may be added as `DescriptiveMetadata/Property[@FormalName="Version"]/@Value`. It would appear only in the case of a version number higher than '1'.
- Subject codes may be added as a set of `DescriptiveMetadata/SubjectCode`.
- The location of origin of the news may be added as `DescriptiveMetadata/Location` (along with an upgrade to NewsML1.2).
- Keywords may also be added as a set of `DescriptiveMetadata/Property[@FormalName="Keyword"]/@Value`.

4.5.2.- Enriched text structure

A multimedia document may include one or more textual components, depending on the product; if multiple textual components are present, each one gets a specific role in the document.

AFP has chosen NITF (an IPTC standard) to represent the structure and content of news articles; many NITF tags are based on HTML4.0 / XHTML. A subset of the NITF structure is included in our NewsML implementation; we chose elements relative to structure only, and avoided those related to the layout of text.

The first **NewsComponent** elements describe the text part

1. NewsComponent/Role/@FormalName: optional; component role (e.g. 'Biography'),
2. NewsComponent/ContentItem: content; see below,

The **ContentItem** element describes the textual content:

3. ContentItem/MediaType/@FormalName: content type ('Text'),
4. ContentItem/Format/@FormalName: content format ('bcNITF2.5'),



5. ContentItem/DataContent: textual content; see below,

Note : AFP only uses the 'body.content' part of NITF to describe enriched text; so the format value ('bcNITF2.5').

The **DataContent** element contains enriched text:

1. DataContent/media: placeholder for an illustration,
2. DataContent/p: paragraph; can contain '**a**' (hyperlinks), '**org**' (organization codes), '**q**' (quotes) tags and '**em**' (emphasis) elements,
3. DataContent/hl2: inter-title,
4. DataContent/ol: ordered list of '**li**' elements,
5. DataContent/ul: unordered list of '**li**' elements,
6. DataContent/dl: HTML-like definition list; contains term/definition pairs ('**dt**', '**dd**').
7. DataContent/table: HTML-like table ('**tbody**', '**tr**', '**td**'),
8. DataContent/pre: preformatted string; spaces are preserved; usually a table-like text.

4.5.2.1.- Illustration placeholders

The <media> element shows the recommended location for an illustration; such a location is set between two paragraphs. The illustration is described in details in a <NewsComponent> element, outside the text. The 'data-location' attribute identifies the component (as an XML fragment identifier). The component gets a local id, and contains the object metadata, including its URL.

1. DataContent/media/@media-type: illustration type ('image' | 'audio' | 'video' | 'other'),
2. DataContent/media/@style: style of the illustration, position along the text ('leftSide' | 'rightSide'),
3. DataContent/media/media-metadata/@name: optional metadata name ; used for a clickable image in order to create a hyperlink (value 'hyperlink'),
4. DataContent/media/media-metadata/@value: optional metadata value ; used for a clickable image, with URL target as value,
5. DataContent/media/media-reference/@alternate-text: picture description, contextual to its use in the text, or description of the hyperlink in the case of a clickable image,
6. DataContent/media/media-reference/@data-location: pointer to the component (*fragment identifier*, beginning with a '#'),
7. DataContent/media/media-reference/@mime-type: mime type of the illustration (this mandatory attribute is empty),

Note: The number of <media> tags may be less than the number of illustration NewsComponents contained with the NewsItem. This happens when a client has not subscribed to a product option like the animated graphics or video clips. In such a case a dummy illustration NewsComponent is still present in the NewsItem, with .gif or .jpg thumbnail and preview plus an empty caption, but no reference to the animated graphic or video. But this component is not referenced in the text as a <media> tag.

Processing the illustrations: The recipient processor should use the illustration placeholders as a trigger for the processing of illustration NewsComponents (see 2.3 Processing model).

Sample:



```

<NewsComponent>
  <ContentItem>
    <MediaType FormalName="Text"/>
    <Format FormalName="bcNITF2.5"/>
    <DataContent>
      <media media-type="image" style="leftSide">
        <media-reference mime-type="" data-location="#photo0"/>
      </media>
      <p>
        text ...
      </p>
    </DataContent>
  </ContentItem>
</NewsComponent>
<NewsComponent Duid="photo0">
  <ContentItem Href="...jpg">
    <MediaType FormalName="Photo"/>
  </ContentItem>
</NewsComponent>

```

4.5.2.2. - Hyperlinks

Hyperlinks – an `<a>` tag – are present inside a paragraph, and used to link the document to other AFP multimedia documents, Web sites or external content, e.g. video clips.

`<a>` attributes :

- href : target URL
- name : link description
- title : in this implementation, linked document's format
- class : hyperlink class (webSite | event | module | document | video | photo | animation).
- style : hyperlink style.
- content : hyperlink description

4.5.2.2.1. - Web site hyperlink

Hyperlinks of this class point to a Web site page, which content is considered as complementary to the current document. Such links are put in a specific paragraph, at the end of the enriched text; this paragraph contains an ordered list (ol), whose role is specified by the 'class' attribute set to 'webset'.

Each link is then placed in an element of the list (li). It's an 'a' tag which attribute is set to 'web', 'href' points to the site, 'name' gives the site name; the tag content describes the link.

Sample:

```

</p>
<ol class="webSet">
  <li>
    <a class="web" href="url" name="site">label</a>
  </li>
  <li>
    <a class="web" href="url" name="site">label</a>
  </li>
</ol>
</p>

```

Filtering the link: the "web list" tag uses an `` structure, and the textual content is the name of the target web page. The complete tag should simply be deleted



if not used.

4.5.2.2.2. - *Rubric hyperlink*

Hyperlinks of this class point to another rubric, such as an 'Event of the Day' rubric.

The 'class' attribute value is set to 'module'; 'name' contains the folder name of the target rubric and 'href' is the relative URL of the index file. The content of the tag describes the target rubric.

Sample:

```
<p>Xxxx. <a class='module' href='../folder/index.xml' name='folder' /> label </a></p>
```

Filtering the link: the textual content is the name of the target rubric. The complete tag should simply be deleted if not used.

4.5.2.2.3. - *Documentation hyperlink*

Hyperlinks of this class point to a given article of a specific rubric called 'documentation' (found in a specific 'doc' directory).

The 'class' attribute is set to 'document', 'name' is the 'slug' (name) of the target document and 'href' is the relative URL of the document. The tag content describes the document.

Sample:

```
<p>Xxxx <a class="document" href="../doc/slug.xml" name=" slug ">label</a> xxxx</p>
```

Note: the 'doc' directory is on the same level as the other rubrics directories.

Filtering the link: this tag is found in a sentence, inside a paragraph. The functionality is optional, and customers who do not use it should filter the tag, but keep its textual content inline.

- If your site doesn't use the 'documentation' rubric, you should filter the NITF hyperlink, but keep the textual content inline.
- If your site uses the 'documentation' rubric, you should process the NITF hyperlink tag in order to obtain a HTML4 hyperlink tag, by just deleting the 'class' attribute.

4.5.2.2.4. - *Video clip hyperlink*

Some AFP multimedia products optionally contain streamed video clips, accessible in different versions (format & quality) via specific hyperlinks.

The number of links varies from a product to one other: if AFP delivers 3 formats in 2 qualities (currently optimized for 56K and 300K transmissions), six links are given. Those links may point to 'metafiles' (managed by the 'video players') that contain links to the real video files.



The video clips are represented by an 'a' tag; the 'class' attribute is set to 'videoSet' and the 'name' attribute is the descriptive label of the video.

This tag contains a series of other 'a' tags, each of which indicates a version of the clip; 'title' gives the format ('windowsMedia', 'real' or 'quickTime' if several streams are in the same file, or 'windowsMedia-low' (low rate), 'windowsMedia-high' (high rate) etc... if each file contains one stream only) and 'href' is the relative URL of the video file.

In this case, the video thumbnail is represented in a similar way to the video clip, i.e. as an anchor of class 'photo', with the 'href' attribute referencing the image; this image is delivered together with the document, as a NewsComponent of MediaType 'Photo'. The 'style' attribute sets the recommended position of the thumbnail in the page ('leftSide' or 'rightSide') (*modified in revision 2*).

Sample:

```
<NewsComponent>
  <ContentItem>
    <MediaType FormalName="Text"/>
    <Format FormalName="bcNITF2.5"/>
    <DataContent>
      <p>xxxxxx</p>
      <p>
        <a class="videoSet" name="Euro">
          <a class="photo" href="#photo0" style="leftSide"/>
          <a class="video" title="windowsMedia-low" href="../../video/afp/fr/mp-lo/name.wmv"/>
          <a class="video" title="windowsMedia-high" href="../../video/afp/fr/mp-hi/name.wmv"/>
          <a class="video" title="realMedia-low" href="../../video/afp/fr/rm-lo/name.rm"/>
          <a class="video" title="realMedia-high" href="../../video/afp/fr/rm-hi/name.rm"/>
          <a class="video" title="quickTime-low" href="../../video/afp/fr/qt-lo/name.mov"/>
          <a class="video" title="quickTime-high" href="../../video/afp/fr/qt-hi/name.mov"/>
        </a> xxxx</p>
    </DataContent>
  </ContentItem>
</NewsComponent>
<NewsComponent Duid="photo0">
  <ContentItem Href="20010528-1000-01.jpg">
    <MediaType FormalName="Photo"/>
  </ContentItem>
</NewsComponent>
```

Filtering the link: the "videoSet" tag has no textual content, and so it should simply be deleted if not used.

Evolution : at present, video clips are treated as 'external' content ; in the future, they will be managed like photos, with a <media> tag of media-type 'video' giving the recommended position of this illustration, itself described in a specific NewsComponent. This component will include the characteristics of the set of included objects (thumbnail and different files, with format and quality metadata).

4.5.2.2.5. - Web graphic hyperlink

AFP multimedia products may contain optional dynamic graphics, distributed in parallel with the XML documents.

The dynamic graphics are represented by an 'a' tag, the 'class' attribute is set to 'animationSet', and 'name' (optional) is a short description (a more complete description can be set in the associated picture caption).



This tag contains two other 'a' tags: the first represents the thumbnail attached to the graphic: it is of class 'photo' and 'href' points to the attached illustration; the second one represents the graphic; the 'class' attribute is set to 'animation'; 'href' is a relative URL which points to the file; the format is specified by a 'title' attribute of value equal 'flash'.

Sample:

```

<NewsComponent>
  <ContentItem>
    <MediaType FormalName="Text"/>
    <Format FormalName="bcNITF2.5"/>
    <DataContent>
      <p>
        <a class="animationSet" name="Euro">
          <a class="photo" href="#photo0" style="leftSide"/>
          <a class="animation" title="flash" href="../../animation/name/start.swf"/>
        </a> xxxx</p>
      </DataContent>
    </ContentItem>
  </NewsComponent>
  <NewsComponent Duid="photo0">
    <NewsLines/>
    <AdministrativeMetadata/>
    <NewsComponent>
      <Role FormalName="Caption"/>
    ...
  </NewsComponent>
  <NewsComponent>
    <Role FormalName="Thumbnail"/>
    <ContentItem Href="...jpg">
      <MediaType FormalName="Photo"/>
    ...
  </ContentItem>
</NewsComponent>

```

Note: the 'animation' directory is on the same level as the directories containing the other rubrics.

Filtering the link: the "animationSet" tag has no textual content, and so it should simply be deleted if not used.

Evolution : Web graphics are at present treated as 'external' content ; in the future, they will be managed like photos, with a <media> tag of media-type 'image' giving the recommended position of this illustration, itself described in a specific NewsComponent.

4.5.2.3.- Organization codes

Organization codes – the <org> tag – are usually placed in the last paragraph of a text component. The 'org' tag contains the name of the organization, and supports two mandatory attributes, 'idsrc' and 'value' ; it can contain also a set of <alt-code> tags, that add alternate values.

<org> and <alt-code> attributes:

- idsrc: code vocabulary (ISIN or SICOVAM),
- value: organization code.

Sample:



```
<p>
<org idsrc="ISIN" value="FR0000130650" >
  DASSAULT SYSTEMES
<alt-code idsrc="SICOVAM" value="13065" />
</org>
</p>
```

4.5.2.4. - Quotes

Quotes – the `<q>` tag – can be included in a paragraph; the selected fragment can, for example, be represented as a string in italics in a web page.

Sample:

```
<p>At the end of 1999, a new branch of AEVR came into being at the request of members of the
European Parliament: an economic observatory. The object of the observatory is to create tangible
tools enabling analysis of wine producing regions.
<q> "It is a question of having the means for analysis and forecast to provide data on which the
experts can work,"</q>
explains Jean-Marc Petit, Director of the Observatory.
</p>
```

4.5.2.5. - Emphasis

Emphasis on a paragraph fragment – the `` tag – can be included in a paragraph; the selected fragment can, for example, be represented as a string in bold in a web page.

4.5.2.6. - Inter-titles

Inter-titles – the `<h2>` tag – can be inserted between two paragraphs; the selected fragment can, for example, be represented as a string in bold and italics in a web page, with a wider font.

4.5.2.7. - Lists

Ordered lists - `` tag – and unordered lists - `` tag - are implemented in the same way as HTML lists, and can be inserted between two paragraphs. The 'ol' or 'ul' element contains a collection of 'li' elements; each 'li' supports enriched text.

4.5.2.8. - Definition lists

Definition lists - `<dl>` tag –are implemented in the same way as HTML lists, and can be inserted between two paragraphs. The 'dl' element contains a collection of 'dt' / 'dd' pairs; 'dt' stands for definition term, 'dd' stands for 'definition data', and may be repeated inside a 'dl' element. 'dt' supports enriched text, and 'dd' contains enriched text in a text block (warning: the HTML syntax is different).

Sample:

```
<NewsComponent>
  <Role FormalName="Personal Data"/>
  <ContentItem>
    <MediaType FormalName="Text"/>
    <Format FormalName="bcNITF2.5"/>
    <DataContent>
      <h2>Brian Laudrup</h2>
      <dl>
        <dt>Name</dt>
        <dd><block><p>Brian Laudrup</p></block></dd>
      </dl>
      <dl>
        <dt>Country</dt>
        <dd><block><p>Danemark</p></block></dd>
      </dl>
    </DataContent>
  </ContentItem>
</NewsComponent>
```



```

<dl>
  <dt>Date of birth</dt>
  <dd><block><p>22-02-1969</p></block></dd>
</dl>
<dl>
  <dt>Place of birth</dt>
  <dd><block><p>Vienna</p></block></dd>
</dl>
</DataContent>
</ContentItem>
</NewsComponent>

```

4.5.2.9. - Tables

Tables - <table> tag – are implemented in the same way as the HTML tables, except that the presence of the element <tbody> is mandatory (a constraint due to NITF2.5).

Sample:

```

<NewsComponent>
  <ContentItem>
    <MediaType FormalName="Text"/>
    <Format FormalName="bcNITF2.5"/>
    <DataContent>
      <table>
        <tbody>
          <tr>
            <td>01</td>
            <td>Michael Schumacher</td>
            <td>Deutschland</td>
            <td>Ferrari</td>
            <td>1:20,447</td>
          </tr>
          <tr>
            <td>02</td>
            <td>Mika Häkkinen</td>
            <td>Finnland</td>
            <td>McLaren</td>
            <td>1:20,529</td>
          </tr>
          <tr>
            <td>03</td>
            <td>David Coulthard</td>
            <td>Schottland</td>
            <td>McLaren</td>
            <td>1:20,927</td>
          </tr>
        </tbody>
      </table>
    </DataContent>
  </ContentItem>
</NewsComponent>

```

Table layout: the NITF table has no layout attribute (colors, column width ...). The addition of those attribute is done during the XML -> HTML transform, using an XSLT stylesheet or any other transform technology.

Evolution: the <tbody> tag is not mandatory anymore in NITF3.0; in the future, AFP will use a new version of NITF and will not use this element anymore.



4.5.2.10.- Preformatted text

Preformatted text – the <pre> tag – can be inserted between two paragraphs. The spaces found in such a fragment must be preserved in order to keep a clean presentation of the text.

4.5.3.- Structure of an illustration

Each illustration is represented by a composite NewsComponent, with a local identifier:

1. NewsComponent/@Duid: local identifier.
2. NewsComponent/NewsLines: textual information, directly publishable; see below.
3. NewsComponent/AdministrativeMetadata: administrative metadata; see below.
4. NewsComponent/NewsComponent: component (enriched text or illustration), **repeatable**; see below.

NewsLines describe some directly publishable textual information about the illustration:

5. NewsLines/HeadLine: headline,

AdministrativeMetadata define administrative information about the illustration:

6. AdministrativeMetadata/Provider/Party/@FormalName: information provider ('AFP'),
7. AdministrativeMetadata/Creator/Party/@FormalName : name and position of the creator,

Evolution: DescriptiveMetadata, like subject codes, location of origin or keywords, may be added in the future.

Illustration NewsComponent elements describe each piece of content; a caption is represented as a 'Text' content, with a 'Caption' role; illustrations are represented as 'Photo' content with one of the following roles: 'Thumbnail' (small size), 'Quicklook' (medium size) or 'Preview' (full screen):

The first NewsComponent (text) is the caption:

8. NewsComponent/Role/@FormalName: component role ('Caption'),
9. NewsComponent/ContentItem: content of the caption; see below,

The ContentItem element of the caption includes:

10. ContentItem/MediaType/@FormalName: component type ('Text'),
11. ContentItem/Format/@FormalName: component format ('bcNITF2.5'),
12. ContentItem/DataContent/p: literal string.

The other NewsComponents contain:

13. NewsComponent/Role/@FormalName: role ('Thumbnail' | 'Quicklook' | 'Preview'),
14. NewsComponent/ContentItem :content of the illustration, see below,

The ContentItem describes the illustration content:

15. ContentItem/@Href: file name (URL); the file is stored in the same folder as the NewsML file,
16. ContentItem/MediaType/@FormalName: media type (Photo ...),
17. ContentItem/Characteristics/Property[@FormalName="Width"]/@Value: illustration height,
18. ContentItem/Characteristics/Property[@FormalName="Height"]/@Value: illustration width.



Evolution: the format of the content may be added in the future under DescriptiveMetadata/Format.

4.5.4.- Examples

4.5.4.1.- Simple text document

Sample:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<NewsML>
  <Catalog Href="http://www.afp.com/dtd/AFPCatalog.xml"/>
  <NewsEnvelope>
    <DateAndTime>20010607T155918Z</DateAndTime>
  </NewsEnvelope>
  <NewsItem>
    <Identification>
      <NewsIdentifier>
        <ProviderId>afp.com</ProviderId>
        <DateId>20010607</DateId>
        <NewsItemId>010607144425.x6pxrl6k</NewsItemId>
        <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
        <PublicIdentifier>urn:newsml:afp.com:20010607:010607144425.x6pxrl6k:1</PublicIdentifier>
      </NewsIdentifier>
      <NameLabel>stw-agr-ali-jur</NameLabel>
    </Identification>
    <NewsManagement>
      <NewsItemType FormalName="News"/>
      <FirstCreated>20010607T144425Z</FirstCreated>
      <ThisRevisionCreated>20010607T144425Z</ThisRevisionCreated>
      <Status FormalName="Usable"/>
    </NewsManagement>
    <NewsComponent>
      <NewsLines>
        <HeadLine>Philip Morris perd 4% et RJR Reynolds, 3,62% à l'ouverture à Wall
Street</HeadLine>
        <DateLine>NEW YORK, 7 juin (AFP)</DateLine>
        <CopyrightLine>© 2001 AFP</CopyrightLine>
        <SlugLine>stw-agr-ali-jur</SlugLine>
      </NewsLines>
      <AdministrativeMetadata>
        <Provider>
          <Party FormalName="AFP"/>
        </Provider>
      </AdministrativeMetadata>
      <DescriptiveMetadata>
        <Language FormalName="fr"/>
        <OfInterestTo FormalName="mmd--francais--journal_internet--une"/>
      </DescriptiveMetadata>
      <NewsComponent>
        <ContentItem>
          <MediaType FormalName="Text"/>
          <Format FormalName="bcNITF2.5"/>
          <DataContent>
            <p>Les actions de Philip Morris et de RJR Reynolds, un autre fabricant de tabac, étaient
en net recul jeudi matin à Wall Street, un jury ayant condamné la veille Philip Morris à payer 3 milliards de dollars
à un malade atteint d'un cancer.</p>
            <p>Philip Morris, le principal perdant, abandonnait 4% à 48 USD et RJR Reynolds
3,62% à 56,75 dollars peu après l'ouverture. Les échanges sur les titres de Loews, l'autre fabricant américain de
tabac, n'avaient pas encore débuté.</p>
            <p>La décision rendue par un jury de Los Angeles, en Californie, est la plus lourde
jamais prise à l'encontre d'une entreprise de tabac dans un procès individuel.</p>
            <p>
              <org idsrc="" value="">PHILIP MORRIS<alt-code idsrc="ISIN"
value="US7181541076"/>
            </org>
          </DataContent>
        </ContentItem>
      </NewsComponent>
    </NewsComponent>
  </NewsItem>
</NewsML>
```



```

        </p>
        <p>
            <org idsrc="" value="">RJ REYNOLDS TOBACCO HOLDINGS<alt-code
idsrc="ISIN" value="US76182K1051"/>
        </org>
        </p>
    </DataContent>
</ContentItem>
</NewsComponent>
</NewsComponent>
</NewsItem>
</NewsML>

```

4.5.4.2.-

Document with a photo

Sample:

```

<NewsML>
  <Catalog Href="http://www.afp.com/dtd/AFPCatalog.xml"/>
  <NewsEnvelope>
    <DateAndTime>20040123T091600Z</DateAndTime>
  </NewsEnvelope>
  <NewsItem>
    <Identification>
      <NewsIdentifier>
        <ProviderId>afp.com</ProviderId>
        <DateId>20040123</DateId>
        <NewsItemId>040123091620.bzjd667i</NewsItemId>
        <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
        <PublicIdentifier>urn:newsml:afp.com:20040123:040123091620.bzjd667i:1</PublicIdentifier>
      </NewsIdentifier>
      <NameLabel>Britain-royal-Diana</NameLabel>
    </Identification>
    <NewsManagement>
      <NewsItemType FormalName="News"/>
      <FirstCreated>20040123T091620Z</FirstCreated>
      <ThisRevisionCreated>20040123T091620Z</ThisRevisionCreated>
      <Status FormalName="Usable"/>
    </NewsManagement>
    <NewsComponent>
      <NewsLines>
        <HeadLine>Police chief ready to question Prince Charles over Diana death</HeadLine>
        <DateLine>LONDON (AFP)</DateLine>
        <CopyrightLine>© 2004 AFP</CopyrightLine>
        <SlugLine>Britain-royal-Diana 2ndlead</SlugLine>
      </NewsLines>
      <AdministrativeMetadata>
        <Provider>
          <Party FormalName="AFP"/>
        </Provider>
      </AdministrativeMetadata>
      <DescriptiveMetadata>
        <Language FormalName="en"/>
        <OfInterestTo FormalName="mmd--english--webwires--europe--uk"/>
      </DescriptiveMetadata>
      <NewsComponent>
        <ContentItem>
          <MediaType FormalName="Text"/>
          <Format FormalName="bcNITF2.5"/>
          <DataContent>
            <media media-type="image" style="rightSide">
              <media-reference mime-type="" data-location="#photo0"/>
            </media>
            <p>Britain's most senior police officer, probing the death of Princess Diana more than six
years ago in Paris, has said he is prepared to question members of the royal family including Prince Charles,
Diana's former husband, about her death.</p>
            <p>Asked on BBC News 24's HardTalk programme whether his inquiries would involve

```




questioning the royal family, Sir John Stevens, the Commissioner of the Metropolitan Police said: "If that's necessary, we'll do that".

Then, asked in the interview -- which was picked up Friday by Britain's Daily Telegraph and Daily Mail newspapers -- if he would quiz Charles, Stevens replied: "Absolutely".

Police in Britain were called in on January 6 to look into the fatal car crash of Princess Diana as it emerged that she had feared Prince Charles was out to harm her.

Opening the first British probe into the death of "the people's princess", coroner Michael Burgess said he had asked Stevens to delve into unrelenting speculation that the deaths of Diana and her boyfriend Dodi Fayed were more than just an accident.

Diana and Fayed, along with their driver Henri Paul, died on August 31, 1997, when their black Mercedes-Benz limousine rammed into a pillar in an underpass beneath the Pont d'Alma in Paris as they sped toward Fayed's apartment.

Following a two-year investigation, the French authorities concluded that the accident was chiefly the result of Paul driving too fast under the influence of alcohol and prescription drugs -- but conspiracy theories have persisted.

The horrific crash came a year after Diana's divorce from Charles, the eldest son of Queen Elizabeth II and heir to the throne, was decreed final following their increasingly stormy 15-year marriage.

In a front-page exclusive on the day the police probe was announced, the mass circulation Daily Mirror newspaper identified Charles as the person whom Diana feared was conspiring to physically harm her 10 months before the Paris crash.

"This particular phase in my life is the most dangerous -- my husband is planning 'an accident' in my car, brake failure and serious head injury in order to make the path clear for him to marry," scribbled Diana in a note to her butler and confidant Paul Burrell, according to the tabloid.

There was no reaction from Buckingham Palace and Charles' official residence Clarence House to the Daily Mirror report, which was repeated in other news media.

"As a police officer, you don't start off with any theories, you go where the evidence takes you," Stevens said in the interview Thursday.

"You have my word we will look at this and by the time this inquiry has been finished and we've looked at every single part of these allegations, we will know what the truth of the matter is and then we will disclose that to the coroner," Stevens said.

</DataContent>

</ContentItem>

</NewsComponent>

<NewsComponent Duid="photo0">

<NewsLines>

<HeadLine>Prince Charles</HeadLine>

</NewsLines>

<AdministrativeMetadata>

<Provider>

<Party FormalName="AFP/Pool"/>

</Provider>

<Creator>

<Party FormalName="John Giles"/>

</Creator>

</AdministrativeMetadata>

<NewsComponent>

<Role FormalName="Caption"/>

<ContentItem>

<MediaType FormalName="Text"/>

<Format FormalName="bcNITF2.5"/>

<DataContent>

Britain's most senior police officer, probing the death of Princess Diana more than six years ago in Paris, has said he is prepared to question members of the British royal family including Prince Charles, pictured here, Diana's former husband about her death

</DataContent>

</ContentItem>

</NewsComponent>

<NewsComponent>

<Role FormalName="Thumbnail"/>

<ContentItem Href="SGE.DLU62.230104091556.photo00.default-70x100.jpg">

<MediaType FormalName="Photo"/>

<Characteristics>

<Property FormalName="Width" Value="70"/>

<Property FormalName="Height" Value="100"/>

</Characteristics>

</ContentItem>

</NewsComponent>



```
<NewsComponent>
  <Role FormalName="Quicklook"/>
  <ContentItem Href="SGE.DLU62.230104091556.photo00.default-245x162.jpg">
    <MediaType FormalName="Photo"/>
    <Characteristics>
      <Property FormalName="Width" Value="245"/>
      <Property FormalName="Height" Value="162"/>
    </Characteristics>
  </ContentItem>
</NewsComponent>
<NewsComponent>
  <Role FormalName="Preview"/>
  <ContentItem Href="SGE.DLU62.230104091556.photo00.default-384x254.jpg">
    <MediaType FormalName="Photo"/>
    <Characteristics>
      <Property FormalName="Width" Value="384"/>
      <Property FormalName="Height" Value="254"/>
    </Characteristics>
  </ContentItem>
</NewsComponent>
</NewsComponent>
</NewsItem>
</NewsML>
```