
Contents

- [1. Introduction](#)
 - [1.1 Early XML Applications](#)
 - [1.2 XML Today](#)
 - [1.3 Some Case Studies](#)
 - [2. CML](#)
 - [2.1 History](#)
 - [2.2 What is CML?](#)
 - [2.3 The Approach](#)
 - [2.4 Tools](#)
 - [3. MathML](#)
 - [3.1 Early Systems](#)
 - [3.2 MathML: Presentation and Content](#)
 - [3.3 Presentation](#)
 - [3.4 Content](#)
 - [3.5 Summary](#)
 - [4. FIXML](#)
 - [4.1 Introduction](#)
 - [4.2 FIX](#)
 - [4.3 FIXML](#)
 - [5. NewsML](#)
 - [5.1 Introduction](#)
 - [5.2 NewsML](#)
 - [5.3 NITF](#)
- Appendices**
- [A. References](#)
 - [B. The Press Release for NewsML](#)
 - [C. A NewsML Document: the Press release for NewsML](#)
 - [D. A NITF Document](#)
 - [E. A NITF Document transformed to HTML](#)
 - [F. XSLT Transformation from NITF to HTML](#)
 - [G. XML Applications](#)

1. Introduction

- [1.1 Early XML Applications](#)
- [1.2 XML Today](#)
- [1.3 Some Case Studies](#)

1.1 Early XML Applications

One of the oldest XML applications is CML, the Chemical Markup Language. It started initially as an SGML application and the authors then became members of the XML Working Group. Two of the earliest XMP Applications sponsored by W3C were MathML and SMIL. Neither were the most obvious applications. The mathematical markup industry was a niche market and could be regarded as a limited application of XML. On the other hand, SMIL (Synchronized Multimedia Integration Language) was aimed at embedded systems which really did not have or want the ability to specify the application via a DTD and allow it to be modified. CML is also rather different from an application that one might expect in that it is mainly about the numerical values associated with molecules.

1.2 XML Today

These early applications were important to XML in that it showed that it could be used outside the bounds of the text document area for which the SGML community had envisaged its use. As John Bosak has pointed out, XML is as important to machine independent data as Java is to machine independent programming. The wealth of tools available and the ability to mix and match facilities from already defined applications gave XML an unexpected leverage. Thus NewsML benefitted from the early work of NITF and the transaction based applications benefitted from the early experience with SMIL.

1.3 Some Case Studies

The four case studies introduced here represent both generic and industry specific languages. We believe that between them they introduce most of the benefits that can be seen from the emergence of XML. They are also rather different from the more run of the mill document oriented or database driven applications.

2. CML

- [2.1 History](#)
- [2.2 What is CML?](#)
- [2.3 The Approach](#)
- [2.4 Tools](#)

2.1 History

The origins of the Chemical Markup Language (CML) date back to the first World-Wide Web Conference (WWW1) held at CERN in May 1994 when a session on the future of HTML developed into a discussion of how Mathematics and Chemistry might be expressed. In late 1994, Henry Rzepa proposed that the output from molecular orbital programs such as MOPAC with regard to molecules, atoms, bonds and their computed properties should be marked up in SGML. In consequence, Peter Murray-Rust developed a prototype CML browser that could interwork with MOPAC as early as 1995. Soon after, CML was formalised as an SGML DTD in 1996. With the arrival of Java, the JUMBO browser was written by Peter Murray-Rust and widely demonstrated. Peter was a member of the XML Working Group and in 1997 CML became the first XML DTD (in any domain) and a working demonstration was presented at WWW6 (1997). Version 1.0 of the CML specification was formally published in 1999.

2.2 What is CML?

CML does not cover all chemistry. It concentrates on molecules. Molecules are comprised of atoms in a particular arrangement. Molecules are thus discrete entities that are representable by a formula (the numbers of atoms of each kind that make up the molecule and a connection table defining how the atoms are connected. The water molecule, for example, is made up of two hydrogen atoms and one oxygen atom, and has the formula H_2O . The connectivity table defines that each hydrogen is connected to the oxygen. In chemistry parlance the hydrogen is said to be bonded to the oxygen and the connectivity table is a list of the bonds in the molecule. CML also enables reactions between molecules to be described, and macromolecular structures (for example proteins which are sequences of simpler molecules called amino acids). It allows quantities and properties to be attached to molecules, atoms or bonds.

In the crystalline state compounds can be described by the arrangement of molecules within what is called a unit cell, which is then replicated through space (like a pattern in SVG!). Crystals are characterised by the type of the unit cell, the dimensions of the unit cell and the molecular arrangement. This latter factor is characterised by the symmetry of the arrangement in what is known as a space group. The details are beyond the scope of this course, suffice it to say that CML aims to represent such aspects of molecular structure.

A very important tool in modern chemistry is spectroscopy, the interaction between light or other forms of radiation and molecules. A spectrum, such as an infrared spectrum (essentially a measure of the fraction of light absorbed by a molecule as a function of the wavelength of the light in the infrared region) serves as a kind of finger print for the molecule and can be used in the identification of unknown molecules and in analytical chemistry to determine the presence of a particular molecule. Spectra can also be represented in CML.

2.3 The Approach

The basic approach to the representation of molecules is to define the molecule by a connectivity table, or **bondArray**, which defines how the atoms listed in an **atomArray** are connected. Different kinds of bonds are possible, for example, single bonds, double bonds, triple bonds, and attributes are defined to capture this.

The molecule, ethanol, provides a simple example. Ethanol, also known as ethyl alcohol, or ethan-1-ol, has the molecular formula, C_2H_6O . The arrangement of these atoms is captured in the structural formula, CH_3CH_2OH . In words, three hydrogens are bonded to the first carbon. The first carbon atom is bonded to the second carbon atom, to which two hydrogens are also attached together with a hydroxyl group, OH (an oxygen bonded to the carbon atom and to a single hydrogen atom). If we label the individual atoms with unique labels, it is clear that this arrangement, or graph (in the mathematical sense of a structure with vertices and edges joining the vertices) can be represented by a list of atoms and bond array. The three dimensional structure of the molecule can be described by giving the 3D (or 2D for a schematic representation) coordinates of the atoms in some convenient cartesian coordinate system.

The figure below shows part of the atom array and bond array that describe the structure of the ethanol molecule.

```
<document>
<!-- CML document - ethanol - karne - 7/8/00 -->
<!-- file converted from: MDL .mol -->
<cml title="ethanol" id="cml_ethanol_karne"
  xmlns="x-schema:cml_schema_ie_02.xml">
  <molecule title="ethanol" id="mol_ethanol_karne">
    <formula>C2 H6 O</formula>
    <string title="CAS">64-17-5</string>
    <float title="molecular weight">46.07</float>
    <atomArray>
      <atom id="ethanol_karne_a_1">
        <float builtin="x3" units="A">1.0303</float>
        <float builtin="y3" units="A">0.8847</float>
        <float builtin="z3" units="A">0.9763</float>
        <string builtin="elementType">C</string>
      </atom>
      <atom id="ethanol_karne_a_2">
        <float builtin="x3" units="A">1.8847</float>
        <float builtin="y3" units="A">1.9889</float>
        <float builtin="z3" units="A">1.5717</float>
        <string builtin="elementType">C</string>
      </atom>
      <atom id="ethanol_karne_a_3">
        <float builtin="x3" units="A">3.1883</float>
        <float builtin="y3" units="A">1.4807</float>
        <float builtin="z3" units="A">1.7425</float>
        <string builtin="elementType">O</string>
      </atom>
      <atom id="ethanol_karne_a_4">
        <float builtin="x3" units="A">0.0000</float>
        <float builtin="y3" units="A">1.2330</float>
        <float builtin="z3" units="A">0.8324</float>
        <string builtin="elementType">H</string>
      </atom>
      ...
    </atomArray>
    <bondArray>
      <bond id="ethanol_karne_b_1">
        <string builtin="atomRef">ethanol_karne_a_1</string>
        <string builtin="atomRef">ethanol_karne_a_2</string>
        <string builtin="order" convention="MDL">1</string>
      </bond>
      <bond id="ethanol_karne_b_2">
        <string builtin="atomRef">ethanol_karne_a_1</string>
        <string builtin="atomRef">ethanol_karne_a_4</string>
        <string builtin="order" convention="MDL">1</string>
      </bond>
      <bond id="ethanol_karne_b_3">
        <string builtin="atomRef">ethanol_karne_a_1</string>
        <string builtin="atomRef">ethanol_karne_a_5</string>
        <string builtin="order" convention="MDL">1</string>
      </bond>
      ...
    </bondArray>
  </molecule>
  ...
</document>
```

At first sight this representation contains some surprises. The tags **atom**, **atomArray**, **bond** and **bondArray** serve obvious functions and are the kinds of tags we would expect to find in CML given the earlier description of the underlying approach. What is surprising is the way in which atoms and bonds are defined. We might have expected to find tags that identify explicitly the coordinates of atoms, and the end points (atoms) of bonds. Instead we have tags that describe data types **float**, **string**. The meanings of these tags are imparted by attributes, in particular, the **builtin** attribute which takes values such as "atomRef", "x3" etc. The **builtin** attribute describes values that CML "knows about". The rationale for this approach is not known, but the impression is that the language has been defined in this way in order to facilitate extension. New entities can be represented in CML by defining new attribute values. It is not necessary (to a certain extent) to define new element types in order to represent new atomic and molecular properties.

Reactions can be represented by a list of reactants and products, plus other elements that represent the reaction conditions (temperature, solvent, etc.). An example is shown below. The Diels-Alder reaction is a very important way of making six-membered rings! This markup again shows the use of generic elements such as **float** and **list**, and the "title" attribute to define the semantics of the element.

```
<reaction title="Diels-Alder cycloaddition" id="simple_rxn_1"
  convention="stepwise">
  <string title="description">
    Simple example of a A + B -> C reaction.
  </string>
  <float title="yield" units="%">88</float>
  <string title="notes">taken from Vollhardt and Schore</string>
  <list title="reactionStep" id="simple_s_1">
    <string title="description">cycloaddition</string>
    <float title="yield" convention="%">88</float>
    <string title="notes">one step</string>
    <link title="reactant" href="simple_mol_reactant1"
      id="simple_lk_1"/>
    <link title="reactant" href="simple_mol_reactant2"
      id="simple_lk_2"/>
    <link title="reagent" id="simple_lk_3">
      <integer title="index">1</integer>
      <string title="solvent">Acetonitrile</string>
      <string title="temperature" convention="degC">100</string>
      <string title="duration" convention="hours">3</string>
      <string title="notes">reflux</string>
    </link>
    <link title="reagent" id="simple_lk_4">
      <integer title="index">2</integer>
      <string title="notes">workup</string>
    </link>
    <link title="product" href="simple_mol_product"
      id="simple_lk_5"/>
    <!-- also catalyst, intermediate,
      transition state as needed -->
  </list>
</reaction>
```

2.3 The Approach

Available tools include:

- CMLDOM-JS: A Javascript implementation of the main components of CML JUMBO3-JS. A Javascript (in-browser) tool to retrieve and display documents containing CML elements.
- SELFML-JS browser. This (Javascript) tool reads one or more SELFML files and displays them, including the emedded CML describing the compounds
- CMLDOM-J. A complete Java implementation of the CML-DOM, extensible to further refinements of CML, developed in parallel with the OMG project.

- JUMBO3-J. A Java browser for any document containing CML elements including 2D and 3D displays.
- [Chimera](#). Working examples of large CML-based documents and scientific articles which use an XSLT stylesheet component library and applets for viewing.
- OpenScience Projects. The OpenScience project to communally develop chemical software tools includes two which have been early adopters of CML; Jmol and JChemPaint.
- JME Editor: Collaborating with the developer of JME (Java Molecular Editor) to create a CML-aware 2D chemical structure editor.
- JMVS. This is a Java3D-based CML-compliant molecular visualiser.
- JChemDig and JChemAgent. Web-based robots which can traverse a remote Site, identify chemical content based on chemical MIME types and create a CML-based database of these files, including derived metadata.
- [JChemValidate](#). An online resource for converting to and digitally signing CML documents.

The Jumbo browser is capable of transforming the CML document into SVG as can be seen in Figure 2.1.

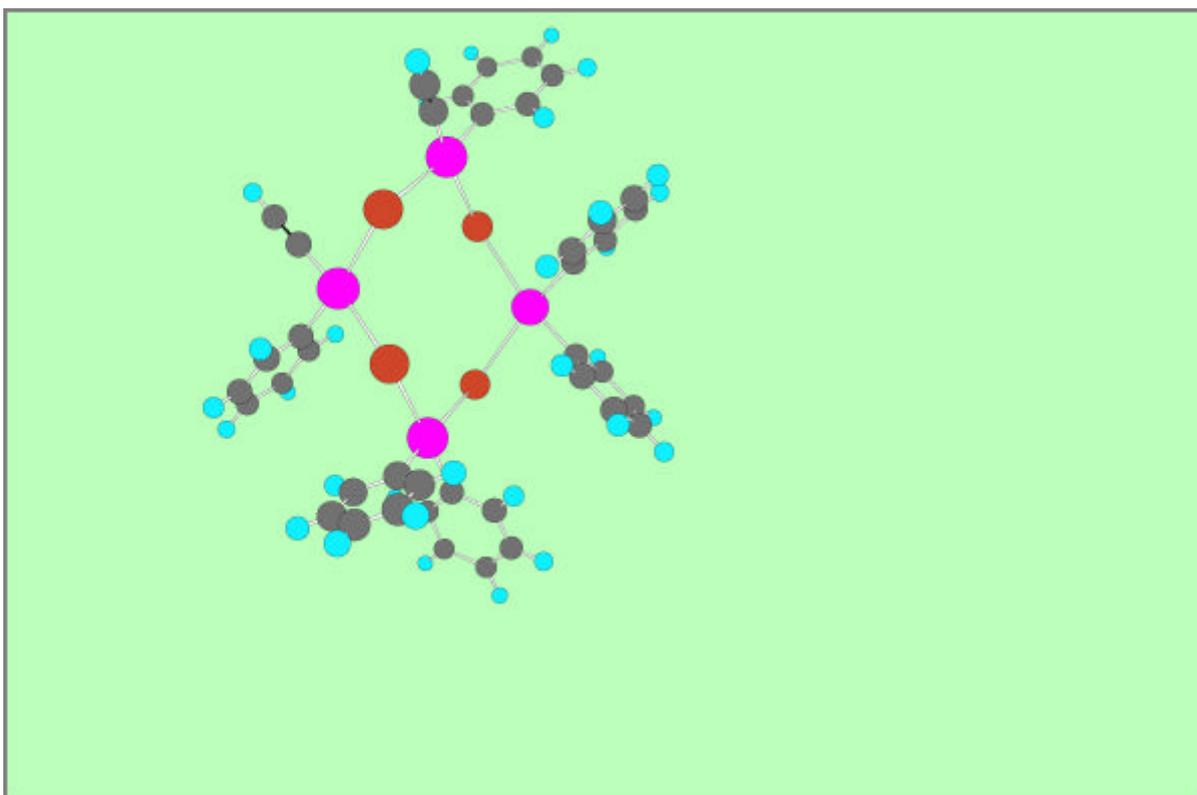


Figure 2.1: CML Transformed to SVG

3. MathML

- [3.1 Early Systems](#)
- [3.2 MathML: Presentation and Content](#)
- [3.3 Presentation](#)
- [3.4 Content](#)
- [3.5 Summary](#)

3.1 Early Systems

Probably the oldest system for expressing and manipulating mathematics on a computer dates back to Tony Hearn's REDUCE system that he developed in the early 1960s at RAND, RAL and Stanford. The first REDUCE Manual was published around 1967. It was implemented in a dialect of Lisp and was widely used in scientific and engineering calculations. There was an earlier extension to FORTRAN by Jean Sammet called FORMAC but that did not have the power to do the mathematics available in REDUCE. MACSYMA was a second system that was widely used early on.

A second seminal system was Formula Algol developed by Al Perlis and Renato Itturiago in the period 1963 to 1970. Formula Algol allowed statements in the language either to be left as algebraic assertions or evaluated as you would in a normal programming language. A presentation system for Formula Algol was developed by Bob Hopgood at Carnegie Tech in 1967.

More recent systems are Steve Wolfram's Mathematica (1988) and Maple from the University of Waterloo in Canada which was started in 1980.

3.2 MathML: Presentation and Content

The W3C Math Working Group was formed in March 1997. MathML is based on the earlier work. As well as the systems described above, it was significantly influenced by Donald Knuth's TeX system, the de facto standard in the mathematical research community for printing mathematics. TeX precisely specifies the positioning of each object that makes up the information to be typeset.

In April 1998, the first version of the Mathematical Markup Language (MathML) was produced with the goal of enabling mathematics to be served, received and processed on the Web. The major innovation was that the two-dimensional symbolic notation was both a definition of the presentation and the content. In consequence, there are three types of elements in MathML:

- **Presentation Elements:** these support the encoding of mathematics for display.
- **Content Elements:** these support the encoding of maths from a semantic point of view.
- **Interface Elements:** these allow a MathML fragment to be embedded in an HTML page.

3.3 Presentation

The MathML presentation model is a hierarchical one similar in some ways to CSS. A MathML expression has a defined rectangular area and more complex expressions are made up from the areas of the simpler expressions that make up the complex expression. For example, to decide how long the line is between the numerator and denominator of a fraction, the length of those two rectangular areas has to be first ascertained and the line is then the longer of the two with the shorter area centred above or below the longer one. The height rather than width will be used to ascertain the height of an integral sign and so on. The presentation markup has the following elements:

mn, mi, ms

Number, identifier, string literal. Numbers normally have an upright font while single character identifiers are normally in italic. However, multi-character identifiers are usually in an upright font.

mo

Operator. Should be displayed as an operator. The presentation and spacing around the operator will depend crucially on which operator it is. The operator element has a set of optional attributes to give hints or be explicit about the presentation required. MathML treats parentheses as operators which mathematically is not true but they are similar as far as presentation is concerned.

mrow

Horizontal group of subexpressions. Normally the contents of an mrow has some semantic meaning such as they form the parts of a subscript.

mfrac

Fraction formed of two subexpressions, the numerator and denominator

msqrt, mroot

Radical formed of subexpressions. msqrt displays the single content child under a square root sign. mroot expects two children where the second is the root to be used.

mfenced

Fences, that is parentheses. Unlike mrow, the parentheses are displayed. Style depends on the attributes.

mstyle

Style settings

mphantom

Used for size calculations

merror

Encloses a syntax error

msub, msup, msubsup, munder, mover, munderover, mmultiscripts

Attach subscripts, superscripts, underscripts and overscripts to a base

mtable, mtr, mtd

Table or matrix, row and element

maction

Creates **live** text in the expression

mspace

Adjustable space

mtext

Arbitrary text

Something simple like $1 + \sin(x)$ would be marked up presentationally as:

```
<math>
  <mrow>
    <mi>1</mi><mo>+</mo><mi>sin</mi>
    <mo>&ApplyFunction;</mo><mo>( </mo><mi>x</mi><mo> )</mo>
  </mrow>
</math>
```

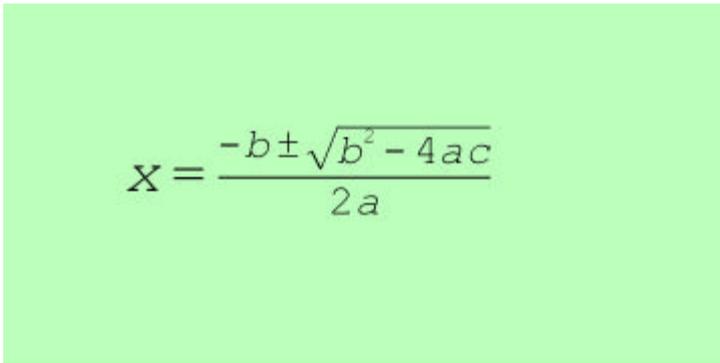
Something like $x^2 + 4x + 4 = 0$ could be marked up as:

```
<mrow>
  <msup>
    <mi>x</mi>
    <mn>2</mn>
  </msup>
  <mo>+</mo>
  <mn>4</mn>
  <mi>x</mi>
  <mo>+</mo>
  <mn>4</mn>
  <mo>=</mo>
  <mn>0</mn>
</mrow>
```

That would display as expected but it really does not bring out the complete structure. In consequence, even in the presentation markup it is more likely that it would be marked up as:

```
<mrow>
  <mrow>
    <msup>
      <mi>x</mi>
      <mn>2</mn>
    </msup>
    <mo>+</mo>
    <mrow>
      <mn>4</mn>
      <mi>x</mi>
    </mrow>
    <mo>+</mo>
    <mn>4</mn>
  </mrow>
  <mo>=</mo>
  <mn>0</mn>
</mrow>
```

A slightly more complex example including its rendering would be:


$$X = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

```
<mrow>
  <mi>x</mi> <mo>=</mo>
  <mfrac>
    <mrow>
      <mrow><mo>-</mo><mi>b</mi></mrow>
      <mo>&PlusMinus;</mo>
      <msqrt>
        <mrow>
          <msup><mi>b</mi><mn>2</mn></msup>
          <mo>-</mo>
          <mrow>
            <mn>4</mn><mo>&InvisibleTimes;</mo>
            <mi>a</mi><mo>&InvisibleTimes;</mo>
            <mi>c</mi>
          </mrow>
        </mrow>
      </msqrt>
    </mrow>
    <mrow>
      <mn>2</mn><mo>&InvisibleTimes;</mo>
      <mi>a</mi>
    </mrow>
  </mfrac>
</mrow>
```

3.4 Content

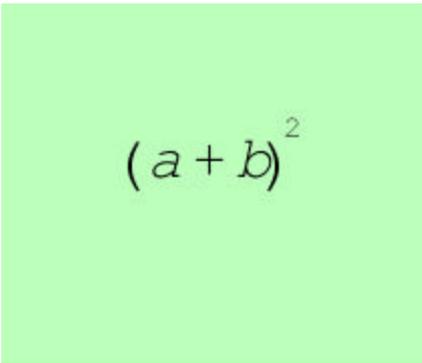
Rather than attempt to define the content of all mathematics, the aim is to capture mathematics needed up to first year university standard. This includes:

- Arithmetic, Algebra and Logic
- Relations
- Calculus
- Set Theory
- Sequences and Series
- Trigonometry
- Statistics
- Linear Algebra
- Semantic Mapping

The difference between the two markups is that an entity like **sin** will have a standard way of presenting it but in terms of content it also has strong semantic relationships with the other trigonometric functions. The expression **1 + sin(x)** would be marked up in terms of content as:

```
<math>
  <apply><plus/><cn>1</cn><apply><sin/><ci>x</ci></apply></apply>
</math>
```

Here is the same formula marked up in both presentation and content markup.



Presentation Markup	Content Markup
<pre><msup> <mfenced> <mrow> <mi>a</mi> <mo>+</mo> <mi>b</mi> </mrow> </mfenced> <mn>2</mn> </msup></pre>	<pre><apply> <power/> <apply> <plus/> <ci>a</ci> <ci>b</ci> </apply> <cn>2</cn> </apply></pre>

4. FIXML

- [4.1 Introduction](#)
- [4.2 FIX](#)
- [4.3 FIXML](#)

4.1 Introduction

Before we discuss FIXML, we need to talk a little bit about the financial markets. Many things get traded on these markets. A few are:

- Shares (sometimes called equities)
- Bonds
- Options
- Futures
- Currencies

All are characterised by having values that change by the millisecond and money is made by buying and selling them (this is an elementary introduction!).

4.1.1 Shares

A share is a part of the capital of a company. A shareholder participates in the management of the company and can receive profits and dispose of the net assets of the company. Shares have a value and that value changes dependent on the position of the company and how it is perceived. In simple terms, if people want to buy the company shares, the price goes up and if not they go down. A stock market is just a market for shares and other items.

4.1.2 Bonds

Unlike shares (often called equities) that represent a part of a company, a bond is a way of lending money to an organisation that can be a government, an agency, a corporation etc. A bond usually has a maturity date (when you get the money back that was borrowed) and interest payments linked to the bond that may or may not be a fixed interest.

4.1.3 Options

A put (call) option is the right to sell (buy) a specified amount of an asset (real or financial) at a fixed price on or before a fixed date. When the holder (investor) acts upon his right to buy (call) or sell (put), the holder **exercises the option**. The price paid for the option is called the **premium**. In Europe, it is normal for the option to be exercised only at the expiry date and not before.

4.1.4 Futures

A futures contract is an agreement to buy or sell an asset at a specific date in the future for a fixed price. That does not mean you have to have the asset or want to buy it but you have the ability to do those things. When the day of reckoning comes you may be lucky or unlucky. So you might agree to buy X at a price of Y in the future. When the day arrives, X might be worth 2Y in which case you make a lot of money as you can immediately sell it. Unlike options, a futures contract is an obligation. The **fixed price** agreed to exchange the underlying asset is known as the **futures price**. The asset is can be a commodity, a bond, a currency, or even an interest rate. These futures get traded all over the world. Some are:

- The Chicago Board of Trade (CBOT).
- The New York Futures Exchange (NYFE).
- The London International Financial Futures Exchange (LIFFE).
- The Swiss Options and Financial Futures Exchange (SOFFEX).
- The European Options Exchange (EOE).
- The Hong Kong Futures Exchange (HKFE).
- The Tokyo International Financial Futures Exchange (TIFFE).

4.1.5 Currency

Currency can be bought and sold. If you are lucky, the exchange rate has moved in your favour between when you purchase the currency and when you sell it. Currency is an asset so there is a futures market and an options market in currency just like other commodities.

4.2 FIX

4.2.1 Introduction

All of the markets above need a standard way of communicating trading information electronically between brokers, buyers and markets. It needs to be flexible given the complexity of the markets and in the early days it needed to be platform independent due to the diversity of systems around.

The Financial Information eXchange (FIX) standard was started in 1993 with a Pilot implementation. By 1995, the standard was defined and in use and developments have continued ever since:

- 1995: FIX 2.7
- 1995: FIX 3.0
- 1997: FIX 4.0
- 2000: FIX 4.2

A typical FIX User is American Century, a large Mutual Fund company in the USA which manages over \$100 Billion in assets. It has used FIX since 1996 as its main trading mechanism. It deals with 65 brokers worldwide, 24 hours a day, dealing in equities, bonds, futures and currencies using FIX. It will handle up to 4 million Indications of Interest (IOIs) a year and the number of FIX transactions is of the order of 40,000 a day.

4.2.2 FIX Protocol

The FIX Protocol consists of:

- A Standard Header
 - BeginString
 - BodyLength
 - MsgType
- A set of Tag=Value fields separated by field delimiters
 - In any order
 - Field Delimiter is <SOH>,
- A Standard Trailer

The FIX Protocol expects the underlying network to deliver messages without failure and in order. However, it does add a sequence number to each message in a session and the individual messages have checksums so a FIX session can detect errors and ask for resends.

Using the symbol ; as a separator, the format of a FIX message looks something like:

```
35=D;55=0001.HK;54=2;38=1000;40=1
```

The tags are numeric values so it is not very readable. Turning it into pseudo English yields:

```
Symbol=0001.HK;Side=Sell;OrderQty=1000;OrdType="Market"
```

This is a person who has 1000 shares of Cheung Kong to sell.


```
<FIXML> <FIXML Message>
<Header>. . .</Header>
<ApplicationMessage>
  <Order>
    <CIOrdID>ORD_1</CIOrdID>
    <HandInst Value="2" />
    <MinQty>1000</MinQty>
    <Instrument>
      <Symbol>
        <IDSource>1</IDSource>
        <SecurityID>277461109</SecurityID>
      </Instrument>
      <Side Value="1" />
      <TransactTime>20000907.09:25:56</TransactTime>
      <OrderQuantity>
        <OrderQty>5000</OrderQty>
      </OrderQuantity>
      <OrderType>
        <LimitOrder Value="2">
          <Price>62.5</Price>
        </LimitOrder>
      </OrderType>
      <Currency Value="usd" />
      <Rule80A Value="A" />
    </Order>
  </ApplicationMessage>
</FIXMLMessage> </FIXML>
```

It is clearly more verbose and that is a worry in this industry sector. Some of the structure that came out in the parsing is now much more obvious and as a result, much easier to process. XML Schema and DTDs can be used to do stronger validation of messages but that is still work in progress. A major advantage is that FIX will be able to interwork with the other E-Commerce applications like SOAP, ebXML and XML Protocol.

The FIXXML DTD at the outer level looks like:

```
<!ELEMENT FIXML (FIXMLMessage+)>
<!ATTLIST FIXML DTDVersion NMTOKEN #FIXED '1.0.0'
  FIXVersion NMTOKEN #FIXED '4.2'
<!ELEMENT FIXMLMessage (Header , ApplicationMessage) >
<!ENTITY % HeaderContent "Sender, OnBehalfOf?, Target,
  DeliverTo?, SendingTime?, PossDupFlag?, PossResend? ">

<!ELEMENT Header (%HeaderContent;)>
<!ELEMENT ApplicationMessage (Advertisement |
  Indication | News | Email | QuoteReq |
  Quote | Order | NewOrderList | ExecutionReport | DK_Trade |
  OrderModificationRequest | OrderCancelRequest | OrderCancelReject |
  OrderStatusRequest | Allocation | AllocationACK |
  SettlementInstructions |
  ListStatus | ListExecute | ListCancelRequest |
  ListStatusRequest | Marketdata | MarketDataInc | MarketDataReq |
  MarketDataReqRej |
  MassQuote | QuoteAck | QuoteCancel |
  QuoteStatusReq | SecurityDef | SecurityDefReq | SecurityStatus |
  TrdSessStatus | TrdSessStatusReq | BusinessReject | Custom ) >
```

The FIXML document looks like:

```
<?xml version='1.0' encoding='UTF-8' ?>
<!DOCTYPE FIXML SYSTEM "fixmlmain.dtd">

<FIXML>
  <FIXMLMessage>
    <Header>
      <Sender>
      </Sender>
      <Target>
      </Target>
      <SendingTime />
    </Header>
    <ApplicationMessage>
    </ApplicationMessage>
  </FIXMLMessage>
</FIXML>
```

An example of a complete message is:

```
<FIXML>
<FIXMLMessage>
<Header>
<Sender> <CompID>Hopgood</CompID> </Sender>
<Target> <CompID>Lloyds</CompID> </Target>
</Header>
<ApplicationMessage>
<Indication>
<IOIid>41926</IOIid>
<Instrument>
<Security> <Symbol>IBM</Symbol> </Security>
</Instrument>
<IOISide Value="1"/>
<IOIShares>2000</IOIShares>
<Price>30.00</Price>
<Currency Value="GBP"/>
<ValidUntilTime>22:50</ValidUntilTime>
</Indication>
</ApplicationMessage>
</FIXMLMessage>
</FIXML>
```

The fields have the following meanings

IOIid

Unique identifier of IOI message.

Instrument

Activity for managing risk (interest rate swaps, etc)

Symbol

Security Symbol

IOISide

Side of Indication (1=Buy, 2=Sell, 7=Undisclosed)

IOIShares

Number of shares in numeric or relative size S=Small,L=Large).

Price

Price per share

Currency

Identifies currency used for price (GBP=£,USD=\$, EUR=euro, CHF= swiss franc etc)

ValidUntilTime

Indicates expiration time of indication message (always expressed in UTC (Universal Time Coordinated))

The FIXML usage is one of receiving messages, massaging the information, onward routing new messages, collecting responses and responding to the original message. This can be error prone and the use of XSLT transformations to apply standard procedures is a real opportunity for the industry. The current FIX network has home built or proprietary FIX engines that need to handle error control and the transmission of messages over a wide range of infrastructures. Again, being able to work over the emerging XML networking facilities will much ease the load on this particular industry and allow the industry to interwork with others. FIX is not the only financial information exchange system. There are other systems for different market areas and interworking is essential if a good image is to be provided to the customer.

5. NewsML

- [5.1 Introduction](#)
- [5.2 NewsML](#)
- [5.3 NITF](#)

5.1 Introduction

News is big business. How it is provided to the individual is many and varied. In consequence, one of the earliest users of the Web was the News industry. Most newspapers now have a Web-based version of their publication. Television companies provide additional material via their web sites. Press Releases are channeled through the Web as much as they are through other media. News generates revenue both through the people who buy the newspapers and the advertisers who use the news vehicles as a way to reach the public.

Selling news is big business and a major characteristic of news is that you can sell it for more than one use. If you have the latest weather forecast, it can be sold to farmers, holiday makers, news channels, television, bookies, the entertainment industry and so on. So if you have a news story, it is valuable to know who might buy it, who bought something similar, what they bought it for etc. So news is not just the story but all the information around the story. Whereas the story itself could probably be marked up in HTML, the complete information surrounding the story requires a much richer environment.

A big player in the news industry is Reuters and they are also a big player in the XMLising of news. But there are other players. The main ones are:

- **International Press Telecommunications Council (IPTC) [6]**: this is the organisation that has overall responsibility for the two XML applications for News: NITF [7] and NewsML[10].
 - IPTC is also responsible for the IPTC/NAA Subject Codes that describe the content of news material.
- **Newspaper Association of America (NAA)**: together with IPTC, NAA developed the News Industry Text Format (NITF) originally in SGML. This was developed to supercede some earlier binary wire formats that were aimed specifically at printed media. NITF is primarily concerned with marking up the news story itself.
- **Publishing Requirements for Industry Standard Metadata (PRISM)**: this has a wider remit than NITF and NewsML and is aiming to produce an XML metadata vocabulary for the catalogue, journal, magazine and news industries. Whereas NewsML and NITF are primarily aimed at the big players in the news industry, PRISM is aimed at a much wider audience.
- **XMLNews**: David Megginson designed a two-part standard [8] called XMLNews-Story and XMLNews-Meta in 1998 that has been widely used on the Web. XMLNews-Story was a subset of NITF. XMLNews-Meta described the news content by giving it a unique identifier, header information, milestones, provenance, rights, subject matter and linking. XMLNews-Meta is expressed in the Resource Description Framework (RDF) and so is part of P08775 and will not be discussed further. David Megginson now works for Reuters on NewsML.

In this primer, we will concentrate on NewsML and NITF, which can be used with NewsML to markup the news story itself. These are not two rival products although one (NITF) is significantly older than the other.

5.2 NewsML

5.2.1 Requirements

NewsML is an XML-based standard to markup and organise news throughout its lifecycle. Good examples of NewsML in use are the IPTC web site [6] and Reuters NewsML showcase [4]. This Primer borrows heavily from the material available at the IPTC site.

The need for NewsML comes from the growth in use and re-use of news throughout the world. This has been partly due to the World Wide Web but also the ability to get news stories to anywhere in the world quickly and the need to fill the cable news channels. The main requirements identified were:

- **News Markup:** not just the new story itself but also the relationship between news items. The photos may arrive separate from the text and archival material may need to be included.
- **Reuse:** the information associated with the story must be such that it can be archived and reused possibly in different contexts.
- **Multimedia:** particularly for the non-paper industry, text, images, video, voice, music all need to be integrated into a single story. The same news story may exist in different formats and be delivered in other formats.
- **Compatible:** it must interwork with existing formats such as NITF even if the long term goal is to replace it.
- **Flexible delivery:** it must be possible to receive all stories or be selective even to the extent of getting only sub parts such as the heading.
- **Development:** news stories occur over a period so the need to update and relate to earlier versions is important.
- **Flexible:** the industry is changing so any standard needs to be future proofed and allow for customisation.
- **Authentication:** it must be possible to gauge the correctness of a story.
- **Efficient:** some communication systems have low bandwidth (often where news is happening) so good use of the bandwidth available is important.
- **Push and Pull:** the need to be able to get the news automatically or only as required.
- **XML:** of course it has to be an XML application to benefit from other activities in related areas.

5.2.2 Structure

The requirements lead to a layered structure that is quite complex but with many optional features so that simple uses can be relatively simple. The NewsML document consists of a set of top-level elements which each contain sub-elements and there is a wide use of attributes to qualify the elements. Figure 5.1 shows the overall structure of a NewsML document.

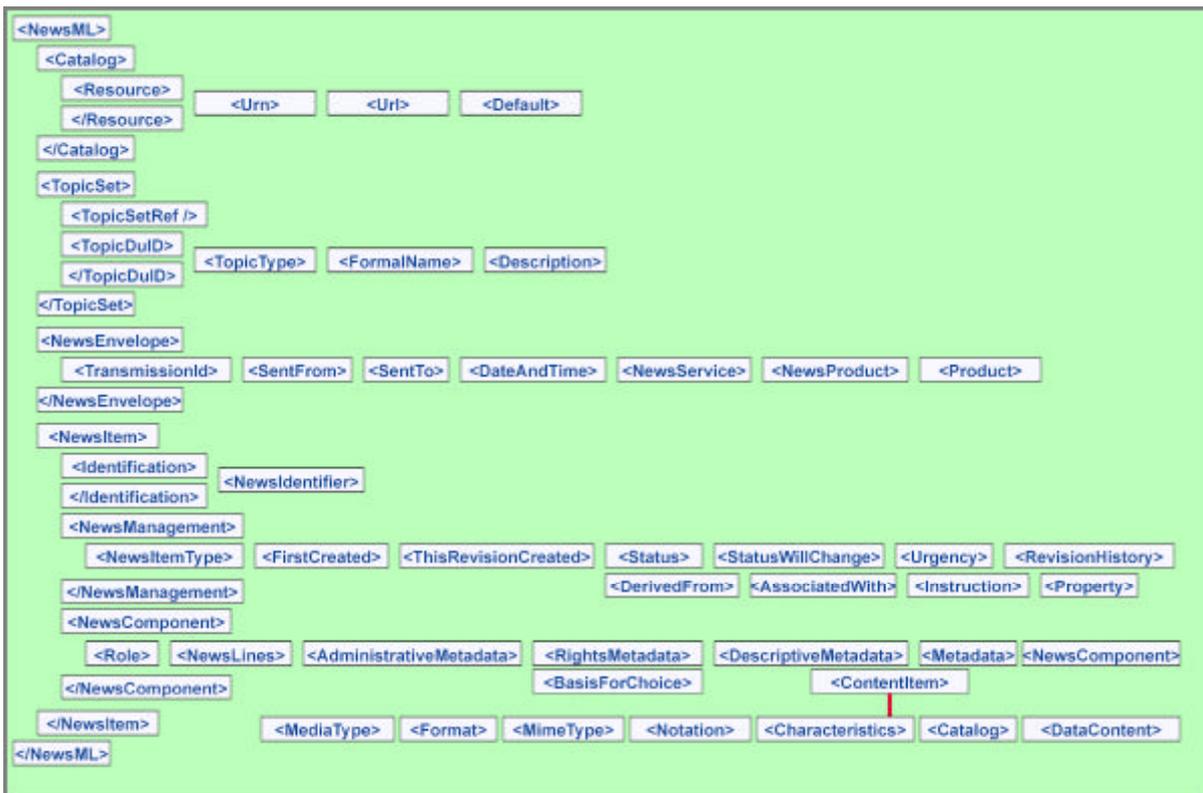


Figure 5.1: NewsML Structure

Down at the bottom right of Figure 5.1 is the [DataContent](#) element that contains the text, image, video etc that is a part of the story. The Mime type for the media is provided as part of the [ContentItem](#) that wraps up the information associated with the data. It is possible to give details of the encoding, the format, its characteristics (file size) etc.

To establish relationships between parts of a story, the [ContentItem](#) is wrapped up in a [NewsComponent](#) which establishes the relationships with other components and adds metadata specific to this component. If the component content is not part of NewsML it is included as the contents of a [NewsLines](#) element (see Figure 5.2). The content of most of the elements inside a [NewsLines](#) element is plain text.

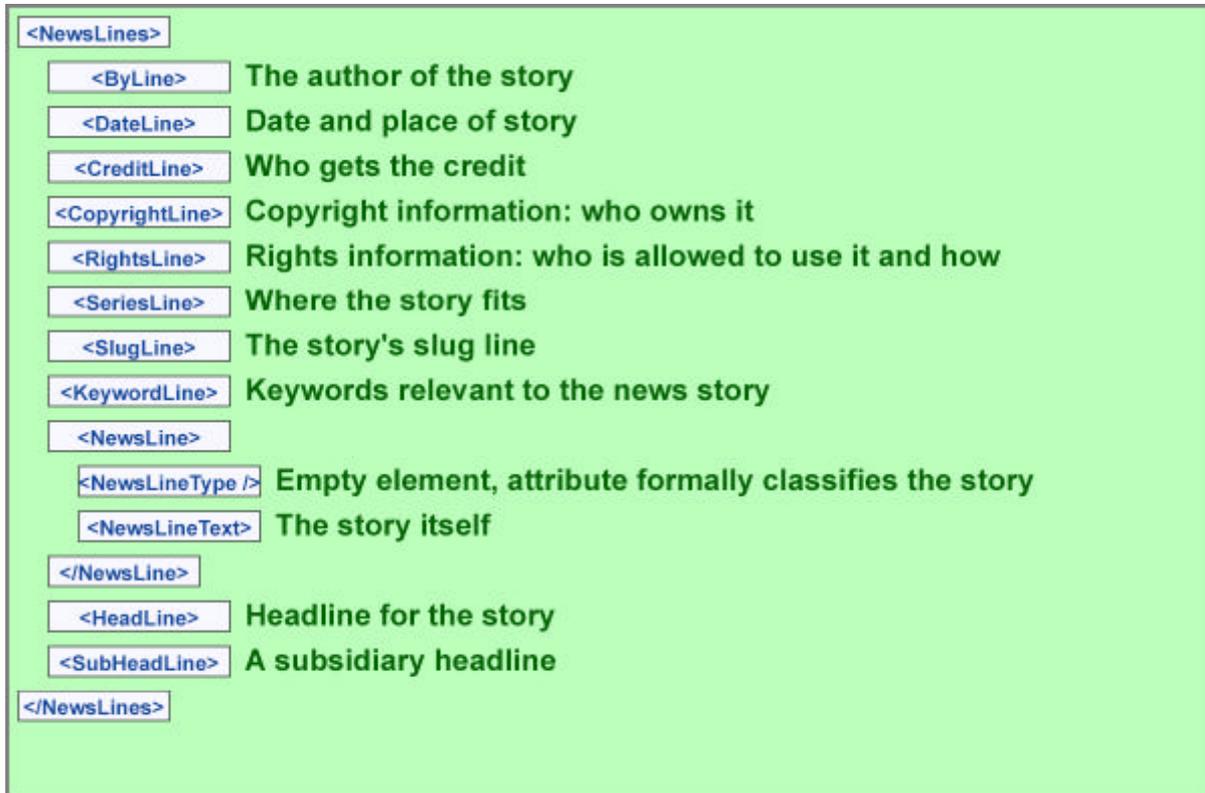


Figure 5.2: NewsLine Element

The element that contains a stand-alone usable piece of news is the [NewsItem](#) element that has a single [NewsComponent](#) element and has sub-elements that describe how you manage the element. One or more [NewsItem](#) elements together make up the story that is the basic NewsML document. As well as the news items themselves, the document contains metadata about the document and workflow related information in a [NewsEnvelope](#) element that says when the story happened, who sent it and where it is to be sent.

The different components in the same document may be different aspects of the story or they may be equivalents. If the story is translated into several languages, they would be included in the same NewsML document. If some statistical information is also available in a graphical form both would be part of the NewsML document. As with SMIL, information is available to make informed decisions as to which format to transmit to a user of the news story.

It is possible for a [NewsComponent](#) to be a news item so there is the ability to construct a large story from smaller stories that exist as news items in their own right. News items can also be used to carry updates to earlier news items to lessen the need for bandwidth. They can also just contain links to the data rather than the data itself. Identifier attributes are used for linking items together and they can either be a [Duid](#) (document-unique identifier) or a [Euid](#) (element-unique identifier).

Appendix B contains a complete NewsML document which is the Press release for NewsML.

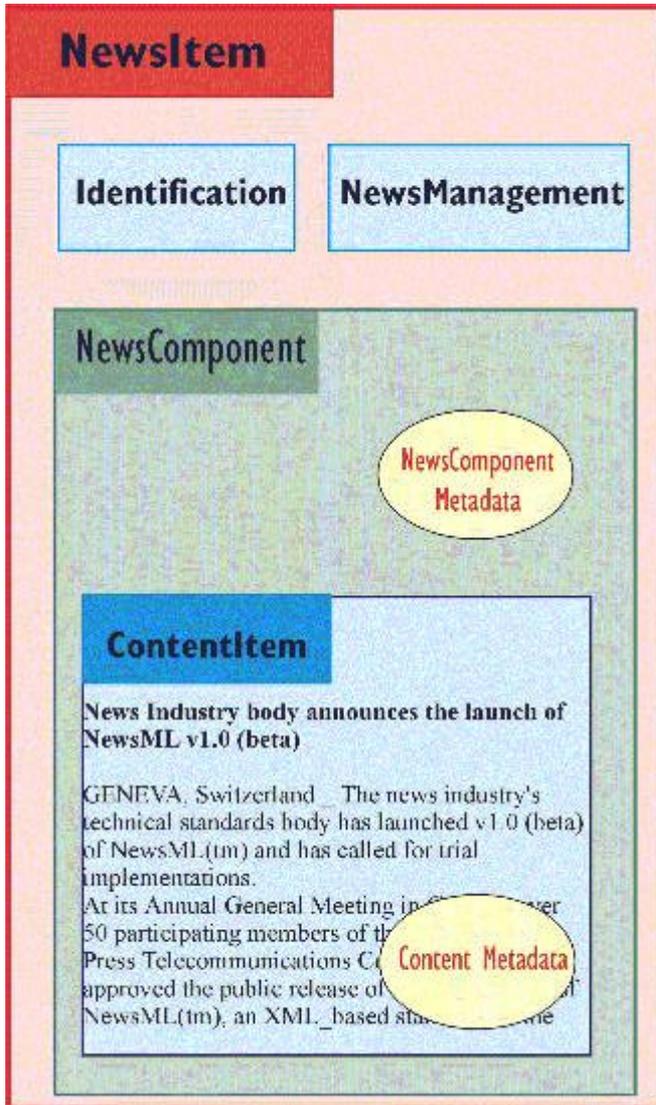


Figure 5.3: NewsItem

Figure 5.3 shows diagrammatically a simple **NewsItem** containing a single news component (an IPTC Press Release on NewsML). Figure 5.4 shows a **NewsItem** consisting of a news component that consists of several news components that make up the story (the development of NewsML). One news component has two equivalent content items (photographs of one of the meetings), one high resolution in colour and one low resolution in black and white. Another has two alternative content items (bottom left) which contain the text of the story in two different languages. On the right is a news component containing a video and another containing the NewsML DTD definition.

NewsItem

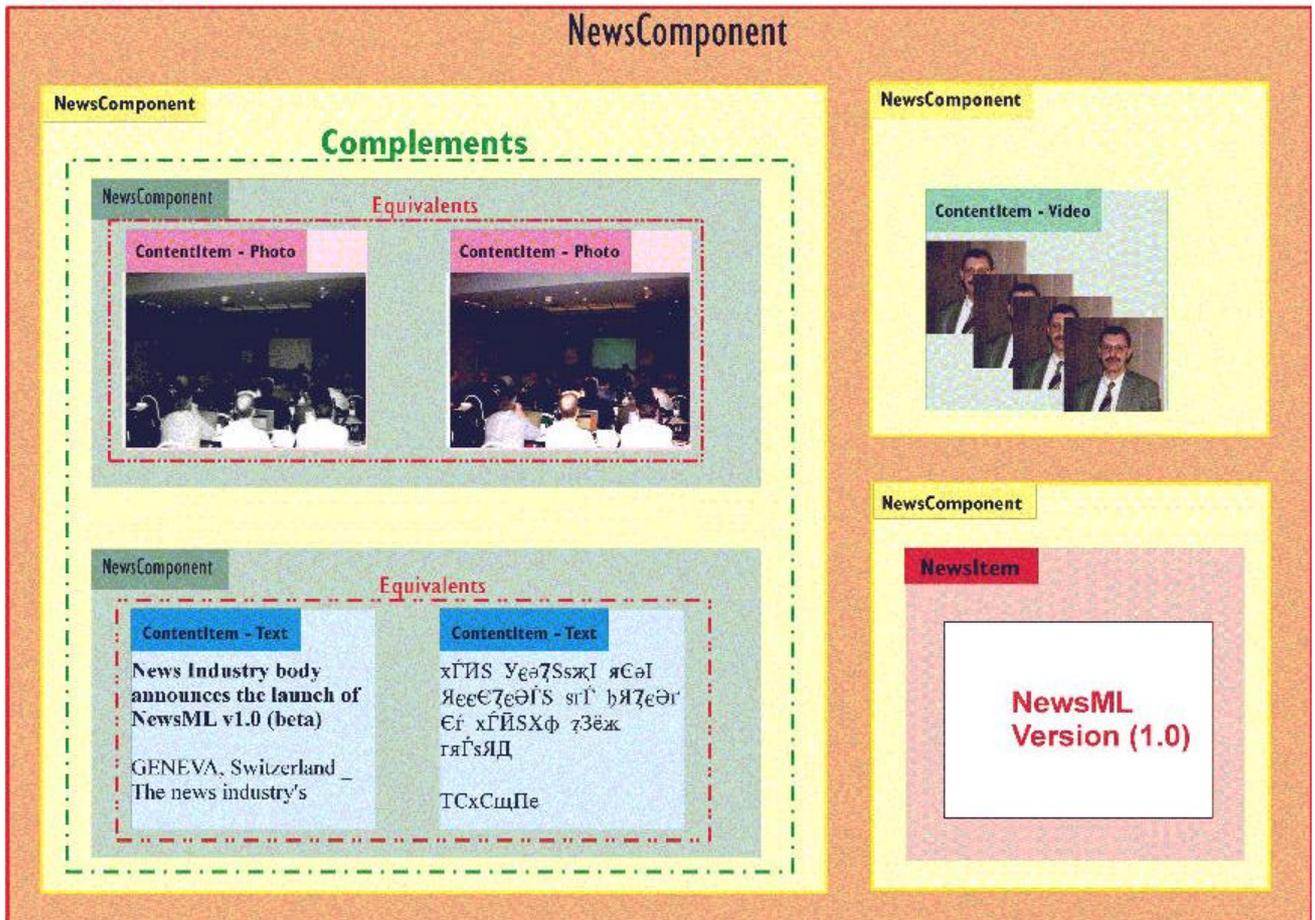


Figure 5.4: Complex NewsItem

Figure 5.5 is similar to Figure 5.1 but emphasises the elements in the [NewsEnvelope](#) and the [NewsItem](#).

A NewsML document has to contain a [NewsEnvelope](#) and at least one [NewsItem](#) which has [Identification](#) and [NewsManagement](#) elements and may contain a [NewsComponent](#) element, and [TopicSet](#) or [Update](#) elements. As we have shown, the [NewsComponent](#) can contain multiple news objects (such as [ContentItems](#)).

An informal meaning of some of the individual elements is as follows:

NewsEnvelope

Information about the transmission of one or more NewsItems as a NewsML document

- **SentFrom**: organisation sending the document
- **SentTo**: organisation to who it is being sent
- **DateAndTime**: date and time in ISO 8601 format
- **NewsService**: identifier defining the news service that owns the [NewsItems](#). This has a formal name defined in a vocabulary
- **NewsProduct**: identifier defining the product that this NewsML document is. Again a formal vocabulary exists for the terms used
- **Priority**: the priority of the NewsML document. Again defined by a formal vocabulary

Catalog

This is a container for resource information. The resources are primarily concerned with metadata to be associated with the [NewsItem](#). Any URN identifies the resource and the URL where to find it.

TopicSet

This is the metadata associated with the item that defines where it is possible to sell it

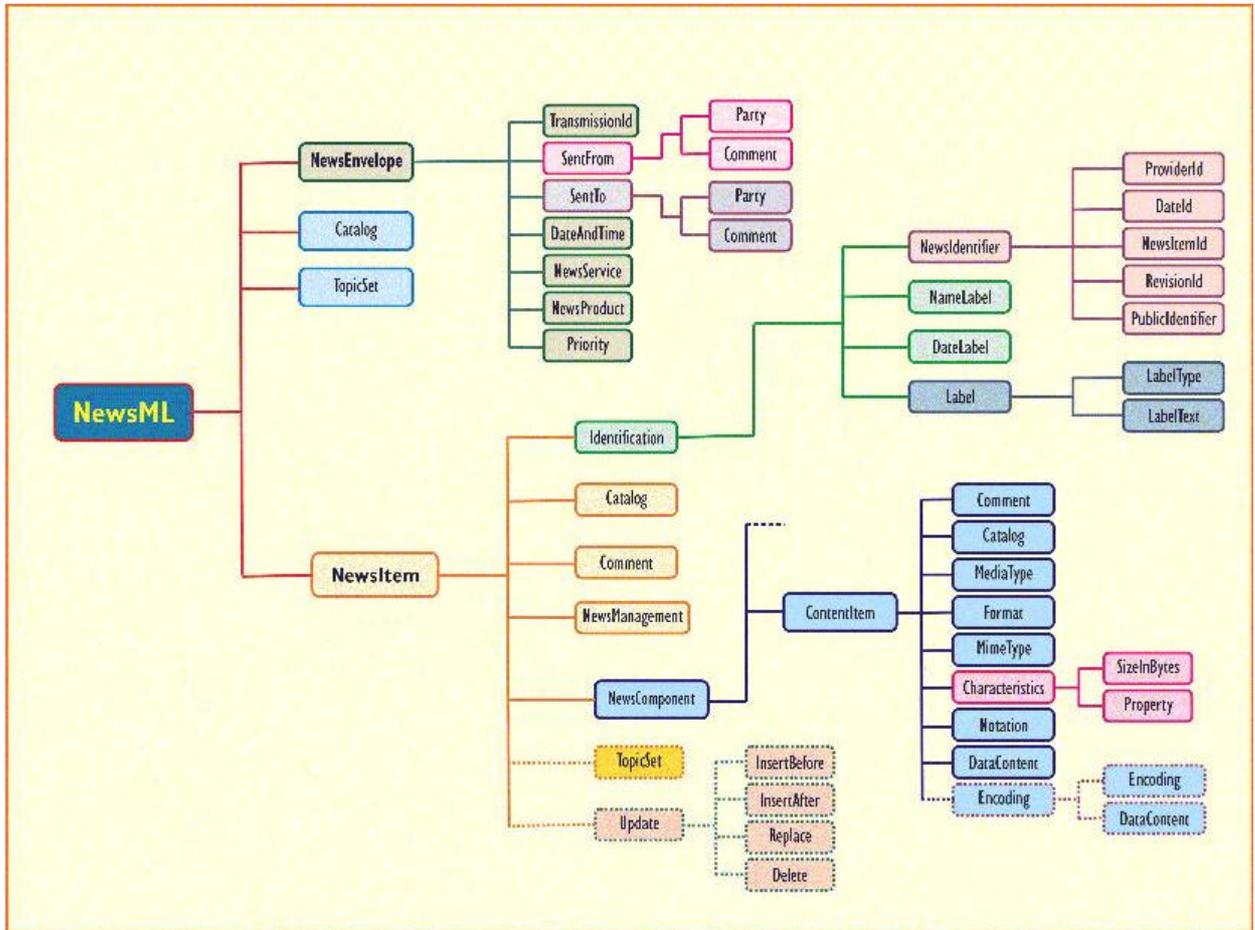


Figure 5.5: NewsML

NewsItem

- **Identification:** this identifies the **NewsItem**
 - **NewsIdentifier:** uniquely identifies the news item
 - **ProviderId:** an internet domain name identifying the provider
 - **DateId:** this is a formal date and time that defines the item throughout any amendments
 - **NewsItemId:** together with the **DateId** it uniquely defines a news item from a provider.
 - **RevisionId:** integer giving the revision number
 - **PublicIdentifier:** a URL that uniquely identifies the item on the Web
- **NewsManagement:** information relevant to managing the news item
 - **NewsItemType:** the type of item defined by a prescribed vocabulary
 - **FirstCreated:** the data and time when it first appeared
 - **ThisRevisionCreated:** time of current revision
 - **Status:** defined against a prescribed vocabulary
 - **StatusWillChange:** defines when its status will change. Used for pressreleases that are not usable until some future time
 - **Urgency:** again a prescribed vocabulary defines the possible values
 - **RevisionHistory:** a pointer to a file containing the history of the item
 - **DerivedFrom:** a reference to another news item to which this one relates
 - **AssociatedWith:** a series of articles or a set of photos would have an appropriate entry
 - **Instruction:** might say that this supercedes a previous one
- **NewsComponent:**
 - **AdministrativeMetadata:** information about who owns the story
 - **RightsMetadata:** information about who is allowed to use it
 - **DescriptiveMetadata:** describes the story (language, relevant genre, subject, who it is of interest to)

What is clear is that a great deal of the information is metadata about the story rather than the story itself.

5.2.3 Metadata

Categorisation of a news item by metadata is a key part of NewsML. NITF has some metadata but it is greatly extended in NewsML. Much of the metadata is optional. The default metadata caters for the normal needs of the news industry but the system has been designed to make enhancements easy. [AdministrativeMetadata](#), [RightsMetadata](#), [DescriptiveMetadata](#) etc have [Property](#) elements that can be used to include additional metadata.

The metadata terms are defined in a set of controlled vocabularies (ontologies or XML schemas) which define allowed values and these are separate from the NewsML DTD. This keeps the syntax separate from the semantics.

Vocabularies are defined as [TopicSets](#). A topic is something like a person, an organisation, a priority, etc. An initial set of [TopicSets](#) is defined for NewsML but additional [TopicSets](#) can be provided. [TopicSets](#) will be part of the metadata description in P08775. The set provided include:

- **Confidence**: the confidence in the story being true
- **Format**: its format
- **Genre**: its characteristics, not its content
- **HowPresent**: how a topic occurs in the content
- **Importance**: significance of the metadata
- **LabelType**: The type of label attached to the item
- **MediaType**: media type
- **MimeType**: the MIME type of the item
- **NewsItemType**: nature of the content
- **Notation**: notation used in the item
- **OfInterestTo**: target audience
- **Priority**: relative importance
- **Property**: named characteristic of an item
- **Provider**: company registered with IPTC and assigned a unique ID
- **Relevance**: how it is relevant to the target audience
- **Role**: distinguishing characteristics of the item
- **Status**: current usability
- **Subject**: description of the content
- **SubjectQualifier**: narrower context such as age or weight of an athlete
- **TopicType**: its type
- **Urgency**: relative importance of the item for editorial examination
- **ISO-Country**: countries listed in the ISO standard for country codes
- **ISO-Currency**: currencies listed in the ISO standard for currency codes
- **ISO-Language**: languages listed in the ISO standard for language codes
- **NAICS-ClassOfIndustry**: class of industry, defined by the North American Industry Classification System
- **NASDAQ-Company**: NASDAQ company code

5.3 NITF

Even though NewsML is media-independent, text is a major part of the content, so specific provisions have been made for text handling. NITF (News Industry Text Format) is the recommended (non-mandatory) format for text markup. Although NITF is quite a rich format, many of the elements are optional so a simple markup can be achieved if necessary. The basic structure is shown in Figure 5.6.

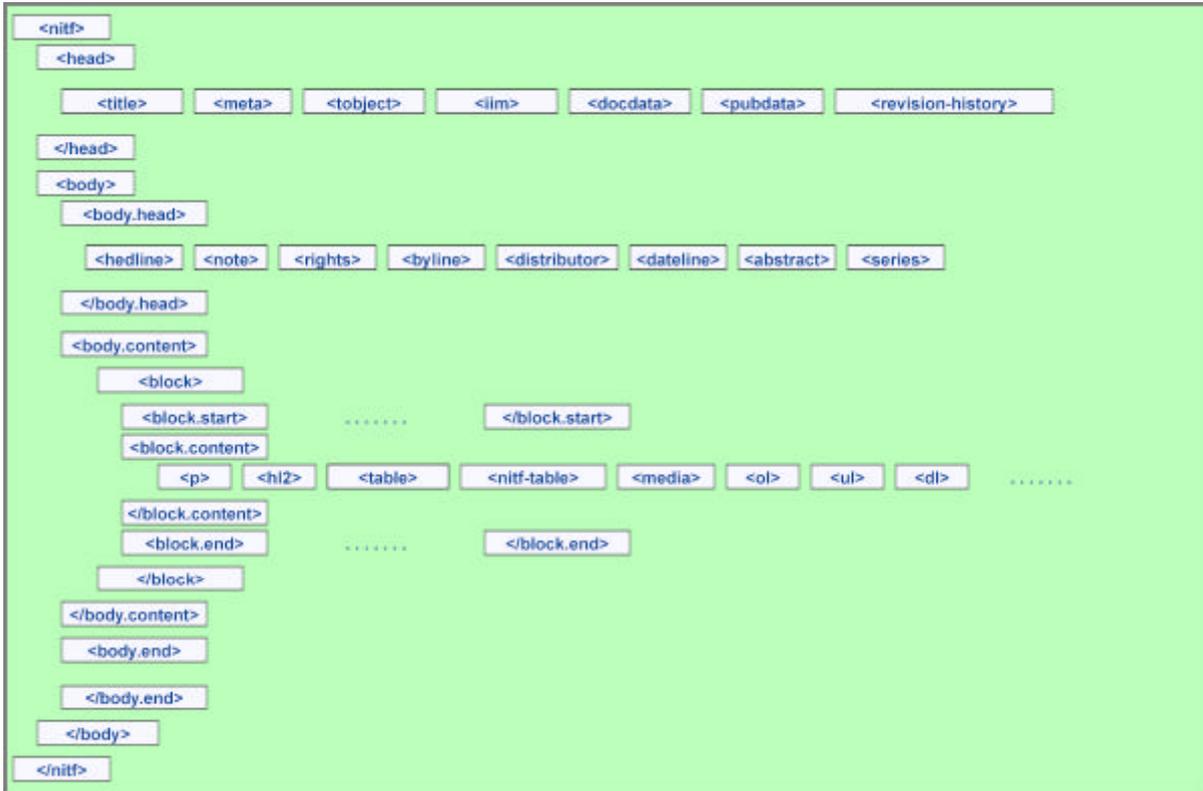


Figure 5.6: Structure of NITF

The head of the document has metadata that in many cases duplicates what is in NewsML. The body of the document has its main content consisting of markup that is very similar to HTML. Appendix D shows an example of a news story (weather forecast) marked up in NITF. This should be compared with Appendix C where the NewsML document has used NITF to mark up the story. In the NewsML example, the metadata is mainly contained in the NewsML markup with NITF just being used for the body of the story while in the NITF document, the metadata is included as part of the NITF document.

Appendix A

References

There are some useful Web sites and books relevant to XML:

1. http://www.xml.org/xmlorg_registry/index.shtml
xml.org's List of XML Applications
2. <http://www.xml-cml.org/>
CML Home Page
3. <http://www.oasis-open.org/cover/>
Robin Cover's XML Cover Page
Addison Wesley, 1998.
4. <http://newsshowcase.reuters.com>
Reuters Showcase site for NewsML
5. <http://www.fixprotocol.org>
FIX Protocol
6. <http://www.iptc.org>
IPTC:
7. <http://www.nitf.org>
NITF:
8. <http://www.xmlnews.org>
XMLNews
9. <http://www.w3.org/TR/xhtml1/>
XHTML Level 1.0 Recommendation, a reformulation of HTML as an XML Application, 26
January 2000
10. <http://www.newsml.org/>
NewsML Web site
11. <http://www.x>
12. <http://www.x>
13. <http://www.x>

Appendix B

The Press Release for NewsML

IPTC membership ratifies NewsML v1.0 and endorses its formal release IPTC PR Committee Amsterdam, October 11, 2000

AMSTERDAM, Netherlands - The news industry's technical standards body has formally ratified v1.0 of its NewsML(TM) standard for the management of multimedia news and announced that it is ready for production use.

At its Autumn Meeting in Amsterdam there was unanimous acceptance amongst the membership that the NewsML v1.0 DTD be formally released having completed a period of beta testing. An updated DTD, functional specification and accompanying examples are now available and a number of members - Agence France-Presse, BusinessWire, Press Association, Reuters, ScreamingMedia, UPI, and Dow Jones'; WSJ.com - have already declared their intention to utilise the new standard.

Klaus Sprick, Senior Vice President of Technology at dpa Deutsche Presse-Agentur and an IPTC Director says, "The membership's endorsement of NewsML v1.0 brings to a close the first phase of the IPTC2000 work programme which we initiated a year ago at a similar meeting in Amsterdam. The goal of IPTC2000 is to deliver an XML-based standard to represent and manage news through its life-cycle, including production, interchange and consumer use. We feel that v1.0 does this and we are pleased to commend this exciting new publishing model to the wider news community".

"NewsML is an extremely powerful and flexible standard, which supports the rich media and multilingual needs of our global client base,"; said Alan Karben, Vice President of Product Development for ScreamingMedia. "Starting this month, we'll be shipping our content publishing system with NewsML built in as a featured multimedia packaging system."

Stuart Myles, Technical Manager at the Wall Street Journal Online, added, "We have tracked the development of NewsML v1.0 very closely and plan to put it in production at WSJ.com".

The IPTC will initiate new work programmes to ensure that NewsML evolves as a standard and achieves widespread acceptance.

What is NewsML?

NewsML is an XML-based standard for all aspects of multimedia news creation, storage and delivery. At the heart of NewsML is the concept of the NewsItem and a NewsItem which can contain various different media - text, photos, graphics, video, - together with all the meta-information that enables the recipient to understand the relationship between components and understand the roles of each component.

Everything the recipient might need to know about the content of the news provided can be included in NewsML's structure. For example, NewsML enables publishers to provide the same text in different languages; a video clip in different formats; or different resolutions of the same photograph. NewsML's rich metadata concept can help with things like revision levels that make it easy to track the evolution of a NewsItem over time, status details (publishable, embargoed, etc.) and administrative details, such as acknowledgements or copyright details. NewsML has default metadata vocabularies to ease implementations but it does not dictate which metadata vocabulary is used (IPTC subject codes, ISO country codes etc.)- a providers just have to indicate which vocabulary they are using. Multiple vocabularies can be utilised within the same NewsItem. For text objects in a NewsItem, the IPTC's News Industry Text Format (NITF) can be utilised.

NewsML is flexible and extensible and uses standard Internet naming conventions for identifying the news objects in a NewsItem. As such, content does not have to actually be embedded within a NewsItem; pointers can be inserted to content held on a publisher's website instead. This means subscribers retrieve the data only when they need to and makes NewsML bandwidth-efficient.

The DTD for NewsML(TM) v1.0, together with a functional specification, supporting documents and background papers can be found on the IPTC web site at <http://www.iptc.org/NewsML> . The DTD is available as a rights-free standard but it remains the intellectual property of the IPTC.

IPTC - Information Technology for News

The International Press Telecommunications Council was established in 1965 to safeguard the telecommunications interests of the World's Press. Since the late 1970's its activities have primarily focussed on developing and publishing Industry Standards for the interchange of news data. At present the IPTC membership is drawn mainly from the major news agencies around the globe but it also has a strong representation from newspaper publishers, system vendors and New Media organisations.

Membership of the IPTC is open to organisations and companies concerned with news collection, distribution and publishing. All existing IPTC standards are copyright IPTC and are administered by the International Press Telecommunications Council, based in England. Information on other IPTC standards such as NITF, IIM and Subject Matter Coding together with a list of existing members is available at <http://www.iptc.org/>.

Companies interested in participating in the IPTC should email David Allen at m_director_iptc@iptc.org. The IPTC is based at Royal Albert House, Sheet Street, Windsor, SL4 1BE . Telephone number +441753705051 or FAX number +441753831541.

NewsML(TM) is a registered trade mark of the IPTC.

<http://www.newsml.org/>

This site is a demonstration showcase for News Markup Language (NewsML; <http://www.xml.com/pub/r/NewsML>), a structure to publish news in any format to any web-enabled device, by the Reuters financial information and news wire service. The "News Page Demo" link takes you to a live, multimedia NewsML demo featuring a current news story. Other sections include technical information and information about NewsML's capability to seamlessly provide news in multiple language formats.

Appendix C

A NewsML Document: the Press Release for NewsML

```
<NewsML>
<Catalog>
  <Resource>
    <Urn> urn:newsml:iptc.org:20001006:topicset.iptc-subjectcode:1 </Urn>
    <Url> ./topicsets/iptc-subjectcode.xml </Url>
    <DefaultVocabularyFor Context = " SubjectDetail " Scheme = " IptcSubjectCodes " ,
    </Resource>
    . . .
  <Resource>
    <Urn> urn:newsml:iptc.org:20020601:MyPropertyTypeTwo:1 </Urn>
    <Url> http://www.mydomain.com/propertytwo.xml </Url>
    <DefaultVocabularyFor
      Context = " Metadata/Property[@FormalName='PropertyTwo']/@Value " />
    </Resource>
</Catalog>

<TopicSet Duid = " newsmltopictypes " FormalName = " TopicType "
  Scheme = " IptcTopicType "> . . .
</TopicSet>

<NewsEnvelope Duid = " nenv01 ">
  <SentFrom>
    <Party FormalName = " IPTC " Scheme = " PubEntities "
      Vocabulary = " #iptc.entities " />
    </SentFrom>
  <SentTo>
    <Party FormalName = " All NewsML Users " Scheme = " PubEntities "
      Vocabulary = " #iptc.entities " />
    </SentTo>
  <DateAndTime> 20020201T120000 </DateAndTime>
  <NewsService FormalName = " IPTC NewsML " Scheme = " PubEntities "
    Vocabulary = " #iptc.entities " />
  <NewsProduct FormalName = " WebSite " Scheme = " PubEntities "
    Vocabulary = " #iptc.entities " />
  <Priority FormalName = " 5 " Scheme = " IptcPriority " />
</NewsEnvelope>

<NewsItem Duid = " FirstHome ">
<Identification Duid = " fhi ">
<NewsIdentifier>
<ProviderId> iptc.org </ProviderId>
<DateId> 20020201 </DateId>
<NewsItemId> WebHome </NewsItemId>
<RevisionId PreviousRevision = " 2 " Update = " N "> 3 </RevisionId>
<PublicIdentifier> urn:newsml:iptc.org:20020201:WebHome:3 </PublicIdentifier>
</NewsIdentifier>
</Identification>
<NewsManagement Duid = " fman ">
<NewsItemType FormalName = " WebContent " Scheme = " IptcWebNewsItemType "
  Vocabulary = " #iptc.webnewsitemtype " />
<FirstCreated> 20020201T120000 </FirstCreated>
<ThisRevisionCreated> 20020203T140000 </ThisRevisionCreated>
<Status FormalName = " Usable " Scheme = " IptcStatus " />
</NewsManagement>
```

```
<NewsComponent Duid = " NMLHome1 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Main " Scheme = " IptcRole " />
<NewsLines Duid = " NLRmt1 ">
<HeadLine> NewsML(TM) in Action </HeadLine>
<SubHeadLine> Intelligent MarkUp for News </SubHeadLine>
<ByLine> David Allen, Managing Director IPTC </ByLine>
<DateLine> 1st February 2002 </DateLine>
<CreditLine> International Press Telecommunications Council </CreditLine>
<CopyrightLine> 2002 (C) IPTC </CopyrightLine>
<RightsLine> May be used without restriction subject to the Licence
    agreement in the NewsML DTD. </RightsLine>
<SeriesLine> No 1 of 1 </SeriesLine>
<SlugLine> NewsML exposed </SlugLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<AdministrativeMetadata Duid = " Admtal ">
<FileName> webpage.xml </FileName>
<SystemIdentifier> webpage.xml </SystemIdentifier>
<Provider>
<Party FormalName = " IPTC " Scheme = " PubEntities "
    Vocabulary = " #iptc.entities " />
</Provider>
<Creator>
<Party FormalName = " David Allen " Scheme = " PubEntities "
    Vocabulary = " #iptc.entities " />
</Creator>
<Source>
<Party FormalName = " IPTC " Scheme = " PubEntities "
    Vocabulary = " #iptc.entities " />
</Source>
<Contributor>
<Party FormalName = " Hugh Johnstone " Scheme = " PubEntities "
    Vocabulary = " #iptc.entities " />
</Contributor>
<Property AllowedValues = " #cres1 " FormalName = " Country "
    Scheme = " MyProperty " Value = " GBR " Vocabulary = " #myprops " />
</AdministrativeMetadata>
<RightsMetadata Duid = " Rtmtal ">
<Copyright>
<CopyrightHolder> IPTC </CopyrightHolder>
<CopyrightDate> 2002 </CopyrightDate>
</Copyright>
<UsageRights>
<UsageType> General </UsageType>
<Geography> Worldwide </Geography>
<RightsHolder> IPTC </RightsHolder>
<Limitations> None </Limitations>
<StartDate> On publication </StartDate>
<EndDate> None </EndDate>
</UsageRights>
</RightsMetadata>
```

```
<DescriptiveMetadata Duid = " desml ">
<Language FormalName = " en " Scheme = " ISO639 " />
<Genre FormalName = " Feature " Scheme = " PubEntities "
  Vocabulary = " #iptc.entities " />
<SubjectCode>
<SubjectDetail FormalName = " 04010006 " Scheme = " IptcSubjectCodes " />
</SubjectCode>
<OfInterestTo FormalName = " General " Scheme = " PubEntities "
  Vocabulary = " #iptc.entities " />
</DescriptiveMetadata>
<Metadata Duid = " next1 ">
<MetadataType FormalName = " PublishingMetadata " Scheme = " MyMeta "
  Vocabulary = " #mymetadata " />
<Property AllowedValues = " #propts " FormalName = " PeriodicalName "
  Scheme = " MyProperty " Value = " IPTCSpectrum#15 " Vocabulary = " #myprops " ,
<Property FormalName = " MediaFormat " Scheme = " MyProperty "
  Value = " LazerPrint " ValueRef = " #proptwol " Vocabulary = " #myprops " />
<Property AllowedValues = " #propts " FormalName = " MediaFormat "
  Scheme = " MyProperty " Value = " PDF " Vocabulary = " #myprops " />
</Metadata>
<NewsComponent Duid = " Compal " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> What is NewsML? </HeadLine>
<ByLine> David Allen, Managing Director IPTC </ByLine>
<DateLine> 1st February 2002 </DateLine>
<CopyrightLine> 2002 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " Reveal ">
<DataContent> An XML-based standard to represent and manage news throughout
its lifecycle, including production, interchange, and consumer use. This web
site is based entirely on NewsML showing how it can form the basis for an
on-line publishing system. The site describes NewsML and allows the detailed
construction of the source files to be examined. </DataContent>
</ContentItem>
</NewsComponent>
```

```
<NewsComponent Duid = " Compa1 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Overview </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " ilc1 ">
<![CDATA[
<Segment>Designed to provide a media-independent, structural framework
for news, NewsML can be applied at all stages in the (electronic) news life
cycle. Typical uses would include: in and between editorial systems; between
news agencies and their customers; between publishers and news aggregators;
and between news service providers and end users. Because it is intended for
use in electronic production, delivery and archiving it does not include
specific provision for traditional paper-based publishing, though formats intended
for this purpose - such as the News Industry Text Format - can be accommodated.
Similarly it is not primarily intended for use in editing or creating news content
though it may be used as a basis for systems doing this.</Segment>
]]>
</ContentItem>
</NewsComponent>
<NewsComponent Duid = " Compa2 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Requirements </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i2c1 ">
<DataContent Duid = " d21 ">
<Article>
<Paragraph>
The need for NewsML comes from the continuing growth in production, use
and re-use of news throughout the world, with rapid expansion of the
Internet being a strong driving force. The set of formal requirements -
given below with explanatory comments -reflect the challenges raised
by these new demands:
<List>
<Item Duid = " reqi1 "> Support the representation of electronic news
entities such as newsitems, parts of newsitems, collections of newsitems,
relationships between newsitems and metadata associated with newsitems.
News may be delivered as single items, or in packages of several related
items, and has to have the metadata to allow efficient production, delivery,
and use (including sorting and searching). </Item>
<Item Duid = " reqi2 "> Be usable throughout the news life cycle. While
the main use will probably be for news interchange, the standard may also
be applied to the creation, management and publication of news in networked
systems, and for archiveapplications. </Item>
<Item Duid = " reqi3 "> Allow newsitems to consist of an arbitrary mixtures
of media types, languages and encodings. News packages can consist different
types of content - text, images, video, audio - all of which are treated
equally. The same newsitem may also exist in a number of different forms,
such as translations of text into different languages or the presentation
of images in alternative formats. </Item>
```

```
<Item Duid =" reqi4 "> Be usable either as a replacement for or allow
the transport of all existing news formats and encodings.The hope is that
NewsML will gradually come to replace older news exchange formats such as
the IIM. However, where other formats perform different functions
(like the NITF with its formatting capabilities) it must be possible to include
them as self-contained items within NewsML. </Item>
<Item Duid =" reqi5 "> Support a number of different physical constructions
of the same data. Depending on user demands, and the delivery systems in use,
there may be a need to supply the same news content in different ways.
Some users may want all of a providers output delivered directly, while others
may prefer to receive notification of availability (with an indication of content
and then retrieve the item if they want to use it. </Item >
<Item Duid =" reqi6 "> Support the management and development of
newsitems overtime. News stories often develop gradually so there is a need
to update, add to, or replace earlier versions. Items in different media
may not be available at the same time, so may have to be brought together. </Item>
<Item Duid =" reqi7 "> Be simply extensible and flexible. Requirements
are liable to change as the markets develop a fixed structure could
rapidly become out-of-date. In addition individual users may wish to add
their own features and extensions. </Item>
<Item Duid =" reqi8 "> Allow for authentication and signature of metadata
and newsitem content.The value of news content, and its associated metadata,
depends on its reliability. </Item>
<Item Duid =" reqi9 "> Not be unduly verbose.Transmission systems
vary in capacity throughout the news industry and the demands on them
keep growing, so there are advantages in keeping the transmission overhead
as small as possible (provided the other requirements are met).
NewsML also needs to be suitable for use with both push and pull delivery system
</Item>
<Item Duid =" reqi10 "> Use XML and other appropriate standards and recommendati:
Adopting XML makes it possible to build on a proven and
fast growing technology and will help to ensure acceptance by the wider
information industry. Since XML is now well established software tools
and development expertise should be generally available. </Item>
</List>
</Paragraph>
</Article>
</DataContent>
</ContentItem>
</NewsComponent>
```

```
<NewsComponent Duid = " Compa5 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Structure </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<NewsComponent Duid = " Compa51 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Description of Structure </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i51c1 " Href = " lStructure.xml ">
<Comment> Intro </Comment>
<MimeType FormalName = " text/xml " Scheme = " IptcMimeType " />
</ContentItem>
</NewsComponent>
<NewsComponent Duid = " Compa52 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Document View </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i52c1 " Href = " images/S_Doctree.jpg ">
<Comment> Document Tree Diagram </Comment>
</ContentItem>
<ContentItem Duid = " i52c2 ">
<Comment> Caption 1 </Comment>
<MimeType FormalName = " text/plain " Scheme = " IptcMimeType " />
<DataContent Duid = " d521 "> A NewsML document has to contain a
NewsEnvelope and at least one NewsItem which has Identification and
NewsManagement elements and may contain a NewsComponent, a TopicSet or
Update elements. The NewsComponent can contain multiple news objects -
such as ContentItems. </DataContent>
</ContentItem>
</NewsComponent>
```

```
<NewsComponent Duid = " Compa53 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Simple NewsItem </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i53c1 " Href = " images/S_SimpleNI.jpg ">
<Comment> Simple NewsItem Diagram </Comment>
<MimeType FormalName = " image/jpeg " Scheme = " IptcMimeTypes " />
</ContentItem>
<ContentItem Duid = " i53c2 ">
<Comment> Caption 2 </Comment>
<MimeType FormalName = " text/plain " Scheme = " IptcMimeTypes " />
<DataContent Duid = " d522 "> Construction of a simple NewsItem
containing a single piece of data - an IPTC Press Release on NewsML.
This item would form part of a basic NewsML element, along with the
NewsEnvelope and, optionally, Catalog and TopicSet elements. </DataContent>
</ContentItem>
</NewsComponent>

<NewsComponent Duid = " Compa54 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supporting " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Complex NewsItem </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i54c1 " Href = " images/S_ComplexNI.jpg ">
<Comment> Complex NewsItem Diagram </Comment>
<MimeType FormalName = " image/jpeg " Scheme = " IptcMimeTypes " />
</ContentItem>
<ContentItem Duid = " i54c2 ">
<Comment> Caption 3 </Comment>
<MimeType FormalName = " text/plain " Scheme = " IptcMimeTypes " />
<DataContent Duid = " d541 "> More complex NewsItem carrying a story
on the development of NewsML. The NewsItem carries a single NewsComponent,
which contains further NewsComponents carrying the news information.
In one there are two equivalent ContentItems carrying different language
versions of the text, another has two equivalent ContentItems carrying
a photograph of one of the meetings. One image is in colour, the other
a reduced definition in black and white. The text and images cover the
same event and so are complements. A separate NewsComponent carries a
related video, while a further NewsComponent carries a NewsItem that has
a recent release of the NewsML DTD - the NewsItem has to be contained
in a ContentItem because surrounding ContentItem can only hold one type
of news object. </DataContent>
</ContentItem>
</NewsComponent>
</NewsComponent>
```

```
<NewsComponent Duid = " Compa3 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> NewsMetadata </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2002 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i3c1 ">
<DataContent Duid = " d31 ">
<nitf baselang = " en.uk " change.date = " 4 July 2000 "
change.time = " 1900 " version = " -//IPTC-NAA//DTD NITF-XML 2.1//EN ">
<head>
<title type = " main "> NewsML </title>
<pubdata date.publication = " 20020201 " />
</head>
<body lang = " en.uk ">
<body.head>
<headline>
<h1> NewsMetadata </h1>
</headline>
<byline> IPTC Staff Writer </byline>
<dateline class = " Windsor October 2000 " />
</body.head>
<body.content>
<block>
<p> Efficient use of metadata is a key feature for NewsML and
considerable effort has been put into the development of a core set of
metadata. This work was able to draw on the substantial intellectual
capital represented by the earlier IIM (Information Interchange Model)
and NITF (News Industry Text Format) standards, but has been substantially
extended, making use of some advanced XML features. </p>

<p> In general, the design of NewsML tries to keep the metadata as
close as possible to the item it describes, while much of the metadata
is optional. In keeping with this, the basic ContentItem has optional
subelements to identify the MediaType (Text, Graphic, Photo, Audio, Video
and Animation); MimeType; Format (such as IIM, DNPR) and Notation (including
SGML, NITF, JPEG and NSK-TIFF). It may also have a characteristics element
to help establish the requirements for the system that has to handle
the data. Examples would include the number of frames for video, or
the duration for audio. The only characteristic directly allowed for
is SizeInBytes but there is a Property element that can be used to specify
other characteristics to meet specific requirements. </p>

<p> NewsComponents are containers that can hold several news objects
of different types (including other NewsComponents), and an essential
feature is the ability to identify the relationships between the objects
and their relative importance. The equivalents list shows which items
are considered to be equivalent to one another, with a BasisForChoice
element identifying information that can be used to choose between the
equivalents. An Essential attribute may be used to show that a given
news object is essential to the meaning of the NewsComponent. Where one
NewsComponent is inside another, the Role attribute specifies its
function (typical roles include Principal, Supporting, Preview, and Abstract).
</p>
```

<p> NewsComponents can also contain AdministrativeMetadata, RightsMetadata, DescriptiveMetadata, and NewsLines. AdministrativeMetadata deals with information about the origin of the NewsItem and includes the file name (along with an optional system address where the item can be found). The Provider and Creator of the news object can be identified, along with the source of the information, while specific provision has been made for identification of syndicated items. A Property element allows for the addition of any other administrative metadata that may be required for specific applications. As the name suggests, RightsMetadata deals with the copyright of the NewsComponent, including details of any usage rights that have been granted to other parties by the copyright holder. Where supplied, this information is in text form along with (optional) links to machine-processable data. </p>

<p> DescriptiveMetadata is used to describe the content of a NewsItem with specific provision made for Language, Genre (the nature of the NewsItem, such as: Current, Analysis, Forecast, Interview, Retrospective); OfInterestTo (target audience), and TopicOccurence. Again, there is a Property element to allow inclusion of any other descriptive metadata needed for a specific application. </p>

<p> NewsLines are used to provide a human-readable version of some of the metadata, and are considered in the NewsText section. </p>

<p> Once the content (within the NewsComponent) has been included in a NewsItem it becomes a piece of "news" and so has to have formal identification and news management features, which are looked at in the NewsManagement section. Provision has also been made for informal identifiers (Labels) to simplify human identification of individual NewsItems. </p>

<p> Although the default metadata has been designed to cater for the routine needs of the news industry it is recognised that many users will want to add their own extensions, and the standard has been specifically designed to make this straightforward. The main metadata categories (AdministrativeMetadata, RightsMetadata, DescriptiveMetadata) have Property elements that can be used for inclusion of additional metadata within the category. There is also a general Metadata element at the NewsComponent level specifically for the addition of new user-defined metadata categories. </p>

<p> Most of the metadata terms are handled as controlled vocabularies - in effect these are lists of allowable values which are maintained separately from the DTD. Having the metadata outside the DTD in this way, greatly simplifies both the general updating and modification of entries, and the development of private metadata sets by users, since there is no need to make changes to the DTD. </p>

<p> The controlled vocabularies are presented as TopicSets, which contain references to individual Topics. In this context, Topics are real things that exist in the outside world. They may be concrete items, such as a person or an organisation, or a more abstract concept, like confidence or priority. In NewsML the Topic elements have a FormalName for identification and a description, along with a Scheme attribute which identifies which particular set of FormalNames is being referred to (this gives positive identification since the same FormalName, on its own, might be used for several Topics). An initial set of TopicSets has been developed for use with NewsML but additional TopicSets can be developed by users to meet their specific needs. </p>

<p> While a major use of TopicSets is as controlled vocabularies for the metadata, the system of TopicSets is also a very powerful tool that makes it possible to link individual NewsML items to the wider context. For example, where reference is made to an individual the TopicSets might include a detailed biography and descriptions of organisations that the person is active in. Main NewsML structural elements can contain a Catalog element which is used to tell the system where it can find resources such as the TopicSets. This is done by using a URN (Uniform Resource Name) and/or one or more URLs (Uniform Resource Locators). The Catalog can also be used to indicate where specific topics appear in the NewsML document (or structural element). </p>

```
</block>
</body.content>
</body>
</nitf>
</DataContent>
</ContentItem>
</NewsComponent>
<NewsComponent Duid = " Compa4 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> With Confidence </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<AdministrativeMetadata>
<FileName> withconfidence.htm </FileName>
</AdministrativeMetadata>
<ContentItem Duid = " i4c1 ">
<MimeType FormalName = " text/html " Scheme = " IptcMimeTypes " />
<Characteristics>
<SizeInBytes> 1760 </SizeInBytes>
</Characteristics>
<Encoding Notation = " Base64 ">
<DataContent Duid = " d41 "> PD94bWwgdmVyc2lvbj0iMS4wIiBlbmNvZGluz0iVVRGLTgiPz41
. . .
0b3BpYy48L3A+DQoJPC9ib2R5Pg0KPC9odGlsPg0K </DataContent>
</Encoding>
</ContentItem>
</NewsComponent>
<NewsComponent Duid = " Compa9 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> NewsManagement </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i9c1 " Href = " NewsManagement.html ">
<Comment> see Story </Comment>
</ContentItem>
</NewsComponent>
```

```
<NewsComponent Duid = " Compa10 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> News Text </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i10c1 " Href = " NewsText.html ">
<Comment> see Story </Comment>
</ContentItem>
</NewsComponent>

<NewsComponent Duid = " Compall1 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> TopicSets </HeadLine>
<ByLine> Hugh Johnstone, Editor IPTC </ByLine>
<DateLine> 1st October 2000 </DateLine>
<CopyrightLine> 2000 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<ContentItem Duid = " i11c1 " Href = " TopicSet.html ">
<Comment> see Story </Comment>
</ContentItem>
</NewsComponent>

<NewsComponent Duid = " Compa6 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> See Markup </HeadLine>
<ByLine> David Allen, MD IPTC </ByLine>
<DateLine> 1st February 2002 </DateLine>
<CopyrightLine> 2002 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<NewsItemRef NewsItem = " lwebpage2.xml ">
<Comment> Reveal Markup </Comment>
</NewsItemRef>
</NewsComponent>

<NewsComponent Duid = " Compa7 " EquivalententsList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> Publishing Options </HeadLine>
<ByLine> David Allen, MD IPTC </ByLine>
<DateLine> 1st February 2002 </DateLine>
<CopyrightLine> 2002 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<NewsItemRef NewsItem = " lwebpage3.xml ">
<Comment> Select Media </Comment>
</NewsItemRef>
</NewsComponent>
```

```
<NewsComponent Duid = " Comp12 " EquivalentList = " no " Essential = " yes ">
<Role FormalName = " Supplementary " Scheme = " IptcRole " />
<NewsLines>
<HeadLine> NewsML Sites </HeadLine>
<ByLine> David Allen, MD IPTC </ByLine>
<DateLine> 1st February 2002 </DateLine>
<CopyrightLine> 2002 (C) IPTC </CopyrightLine>
<KeywordLine> NewsML </KeywordLine>
</NewsLines>
<NewsItemRef NewsItem = " lwebpage6.xml ">
<Comment> Other NewsML Users </Comment>
</NewsItemRef>
</NewsComponent>
</NewsComponent>
</NewsItem>
</NewsML>
```

Appendix D

A NITF Document

```
<nitf>
<head>
  <title>Norfolk Weather and Tide Updates</title>
  <tobject tobject.type="news">
    <tobject.subject
      tobject.subject.refnum="17000000"
      tobject.subject.type="Weather"
    />
    <tobject.subject
      tobject.subject.refnum="04001002"
      tobject.subject.detail="Fishing Industry"
    />
  </tobject>
  <docdata>
    <identified-content>
      <location
        location-code="23602"
        code-source="zipcodes.usps.gov"
      />
    </identified-content>
  </docdata>
</head>
<body>
  <body.head>
    <headline>
      <h1>Weather and Tide Updates for Norfolk</h1>
      <h2>A sample, fictitious NITF article</h2>
    </headline>
    <note><body.content><p>This sample article was created completely
      from scratch in order to illustrate various features of NITF.
      Parts of it are somewhat contrived, in order to illustrate as much
      of the DTD as possible.</p></body.content></note>
    <byline>
      By <person>Alan Karben</person>
      <byttl>NITF Network News Online</byttl>
    </byline>
  </body.head>
```

```
<body.content>
  <p>The weather was great today in Norfolk, Virginia. Made me want to take
  out my boat, manufactured by the <org value="acm" idsrc="iptc.org">
  Acme Boat Company</org>.</p>
  <p>Tides in Norfolk are running normal today. This weeks article highlights
  many of this week's fishing issues, and also presents a reference table
  of tide times.</p>
  <h2>The Tides are High</h2>
  <p>As can be seen from the table below, the shores of Oceanview again
  present the brightest spots for fishermen and sandcastle-builders
  alike.</p>
  <nitf-table>
    <nitf-table-metadata>
      <nitf-table-summary>
        <p>This is a table filled with weather data, good for fishermen
        living in Norfolk, Virginia.</p>
      </nitf-table-summary>
      <nitf-col value="beach"/>
      <nitf-col value="day-high"/>
      <nitf-col value="day-low"/>
      <nitf-col occurrences="2" value="tide-time"/>
      <nitf-colgroup occurrences="3" value="three-day-forecast">
        <nitf-col value="day-high"/>
        <nitf-col value="day-low"/>
      </nitf-colgroup>
    </nitf-table-metadata>

    <table border="1">
      <tr>
        <!-- beach -->
        <th></th>

        <!-- day high and low -->
        <th colspan="2">today</th>

        <!-- tide times -->
        <th colspan="2">tide</th>

        <!-- forecast tomorrow -->
        <th colspan="2">tomorrow</th>

        <!-- forecast the next day -->
        <th colspan="2">next day</th>

        <!-- forecast the day after that -->
        <th colspan="2">third day</th>
      </tr>
```

```
<tr>
  <!-- beach -->
  <th>beach</th>

  <!-- day high and low -->
  <th>high</th>
  <th>low</th>

  <!-- tide times -->
  <th>in</th>
  <th>out</th>

  <!-- forecast tomorrow -->
  <th>high</th>
  <th>low</th>
  <!-- forecast the next day -->
  <th>high</th>
  <th>low</th>

  <!-- forecast the day after that -->
  <th>high</th>
  <th>low</th>
</tr>
<tr>
  <!-- beach -->
  <td>Sunset</td>

  <!-- day high and low -->
  <td>30</td>
  <td>14</td>

  <!-- tide times -->
  <td>09:23</td>
  <td>18:51</td>

  <!-- forecast tomorrow -->
  <td>28</td>
  <td>11</td>

  <!-- forecast the next day -->
  <td>31</td>
  <td>12</td>

  <!-- forecast the day after that -->
  <td>33</td>
  <td>9</td>
</tr>
```

```
<tr>
  <!-- beach -->
  <td>Oceanview</td>

  <!-- day high and low -->
  <td>31</td>
  <td>15</td>

  <!-- tide times -->
  <td>09:25</td>
  <td>18:56</td>

  <!-- forecast tomorrow -->
  <td>26</td>
  <td>11</td>

  <!-- forecast the next day -->
  <td>31</td>
  <td>11</td>

  <!-- forecast the day after that -->
  <td>31</td>
  <td>9</td>
</tr>
<tr>
  <!-- beach -->
  <td>Shellfish</td>

  <!-- day high and low -->
  <td>29</td>
  <td>15</td>

  <!-- tide times -->
  <td>09:25</td>
  <td>18:53</td>

  <!-- forecast tomorrow -->
  <td>26</td>
  <td>9</td>

  <!-- forecast the next day -->
  <td>29</td>
  <td>11</td>

  <!-- forecast the day after that -->
  <td>30</td>
  <td>11</td>
</tr>
</table>
</nitf-table>
```

```
<p>Based on these tide tables, I believe you can see that this
    weekend stands to be an excellent one for small- or large-scale
    fishing exhibitions.</p>
<media media-type="image" style="align:right">
  <media-reference
    mime-type="image/jpeg"
    source="high-tide.jpg"
    alternate-text="The tides are high."
    height="185"
    width="278"
  >
  </media-reference>
  <media-caption>
    The tides, captured on film late yesterday.
  </media-caption>
  <media-producer>
    Karben
  </media-producer>
</media>
<p>There are many local nooks that fishing fans may want to keep a special
    eye on.
</p>
<ul>
<li><em>Deer Creek:</em> This area has proven to be a bass-lover's
    haven. Wilt Monthaven reports that examples of bass over 30 inches
    long have been reeled-in by both the casual and the professional
    angler.</li>
<li><em>Fox Run:</em> A bit more difficult, logistically, to
    navigate. However, the reports of Sturgeon in this area have
    put it on the fishing map for the first time since 1996.</li>
<li><em>Pheasant Hollow:</em> If you don't mind the crowds,
    this old favorite has come through this year. More fish
    than you can shake a stick at. Or a rod at.</li>
</ul>
<p>Happy fishing everybody!</p>
</body.content>
<body.end>
  <tagline>Stewart Klometers contributed to this article.</tagline>
</body.end>
</body>
</nitf>
```

Appendix E

A NITF Document transformed to HTML

Weather and Tide Updates for Norfolk

A sample, fictitious NITF article

Editor's Note: This sample article was created completely from scratch in order to illustrate various features of NITF. Parts of it are somewhat contrived, in order to illustrate as much of the DTD as possible.

By **Alan Karben**

NITF Network News Online

The weather was great today in Norfolk, Virginia. Made me want to take out my boat, manufactured by the [Acme Boat Company](#).

Tides in Norfolk are running normal today. This weeks article highlights many of this week's fishing issues, and also presents a reference table of tide times.

The Tides are High

As can be seen from the table below, the shores of Oceanview again present the brightest spots for fishermen and sandcastle-builders alike.

	today		tide		tomorrow		next day		third day	
beach	high	low	in	out	high	low	high	low	high	low
Sunset	30	14	09:23	18:51	28	11	31	12	33	9
Oceanview	31	15	09:25	18:56	26	11	31	11	31	9
Shellfish	29	15	09:25	18:53	26	9	29	11	30	11

Based on these tide tables, I believe you can see that this weekend stands to be an excellent one for small- or large-scale fishing exhibitions.

There are many local nooks that fishing fans may want to keep a special eye one.

The tides, captured on film late yesterday.

- **Deer Creek:** This area has proven to be a bass-lover's haven. Wilt Monthaven reports that examples of bass over 30 inches long have been reeled-in by both the casual and the professional angler.
- **Fox Run:** A bit more difficult, logistically, to navigate. However, the reports of Sturgeon in this area have put it on the fishing map for the first time since 1996.
- **Pheasant Hollow:** If you don't mind the crowds, this old favorite has come through this year. More fish than you can shake a stick at. Or a rod at.

Happy fishing everybody!

Stewart Klometers contributed to this article.

Appendix F

XSLT Transformation from NITF to HTML

This is the XSLT transformation which when applied to Appendix D results in Appendix E.

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">

<xsl:strip-space elements="*" />

<xsl:output method="html" encoding="ISO-8859-1" />

<!-- Created by Johan Lindgren (TT, Sweden) and Alan Karben (ScreamingMedia, US)
to show various possible outputs from NITF.
It's not intended to handle all possible combinations of data.
-->

<!--      MAIN TEMPLATE      -->
<xsl:template match="/">
<html>
  <head>
    <title><xsl:value-of select="nitf/head/title"/></title>
    <link rel="stylesheet" type="text/css" href="nitf.css"/>
  </head>
  <body><table border="1" cellpadding="6" width="550"><tr><td>
    <xsl:apply-templates />      <!-- Call all subtemplates -->
  </td></tr></table></body>
</html>
</xsl:template>

<xsl:template match="body.head|body.content">
  <xsl:apply-templates />
</xsl:template>

<xsl:template match="p">
  <p class="nitfp"><xsl:apply-templates /></p>
</xsl:template>

<xsl:template match="title">
</xsl:template>

<!-- table -->

<xsl:template match="nitf-table-summary">
</xsl:template>
```

```
<xsl:template match="table">
  <xsl:element name="table">
    <xsl:attribute name="border">
      <xsl:value-of select="@border"/>
    </xsl:attribute>
    <xsl:apply-templates />
  </xsl:element>
</xsl:template>
<xsl:template match="tr">
  <tr><xsl:apply-templates /></tr>
</xsl:template>

<xsl:template match="th">
  <xsl:element name="th">
    <xsl:attribute name="colspan">
      <xsl:value-of select="@colspan"/>
    </xsl:attribute>
    <xsl:apply-templates />
  </xsl:element>
</xsl:template>

<xsl:template match="td">
  <td><xsl:apply-templates /></td>
</xsl:template>

<xsl:template match="byline">
  <p class="nitfby">
    <xsl:apply-templates/>
  </p>
</xsl:template>

<xsl:template match="person">
  <b><xsl:value-of select="."/></b>
</xsl:template>

<xsl:template match="byttl">
  <br/><i><xsl:value-of select="."/></i>
</xsl:template>

<xsl:template match="hedline">
  <div class="hedline"><xsl:apply-templates /></div>
</xsl:template>

<xsl:template match="h1">
  <h1 class="nitfh1"><xsl:apply-templates /></h1>
</xsl:template>

<xsl:template match="h2">
  <h2 class="nitfh2"><xsl:apply-templates /></h2>
</xsl:template>

<xsl:template match="h3">
  <h3 class="nitfh3"><xsl:apply-templates /></h3>
</xsl:template>
```

```
<xsl:template match="note">
  <div class="note"><blockquote><i>Editor's Note:</i>
  <xsl:value-of select="."/;></blockquote></div>
</xsl:template>

<xsl:template match="tagline">
  <p class="tagline"><i><xsl:value-of select="."/;></i></p>
</xsl:template>

<xsl:template match="ul">
  <ul><xsl:apply-templates /></ul>
</xsl:template>

<xsl:template match="li">
  <li><xsl:apply-templates /></li>
</xsl:template>

<xsl:template match="em">
  <b><xsl:apply-templates /></b>
</xsl:template>

<xsl:template match="org">
  <b>
  <xsl:element name="a">
  <xsl:attribute name="href">http://www.stockpoint.com/get-quote?ticker=
  <xsl:value-of select="@value"/></xsl:attribute>
  <xsl:attribute name="class">org</xsl:attribute><xsl:value-of select="."/;>
  </xsl:element>
  </b>
</xsl:template>

<!--
<xsl:template match="media">
  <table border cellpadding="4" align="right">
  <xsl:element name="a">
  <xsl:attribute name="href">http://www.stockpoint.com/get-quote?ticker=
  <xsl:value-of select="@value"/></xsl:attribute>
  <xsl:attribute name="class">org</xsl:attribute><xsl:value-of select="."/;>
  </xsl:element>
  </b>
  </table>
</xsl:template>
-->
```

```
<xsl:template match="media">
  <xsl:element name="table">
    <xsl:attribute name="align">right</xsl:attribute>
    <xsl:attribute name="border">1</xsl:attribute>
    <xsl:attribute name="width"><xsl:value-of
      select="media-reference/@width"/></xsl:attribute>
    <xsl:attribute name="cellpadding">6</xsl:attribute>
    <tr><td>
      <xsl:element name="img">
        <xsl:attribute name="src">images/<xsl:value-of
          select="media-reference/@source"/></xsl:attribute>
        <xsl:attribute name="width"><xsl:value-of
          select="media-reference/@width"/></xsl:attribute>
        <xsl:attribute name="height"><xsl:value-of
          select="media-reference/@height"/></xsl:attribute>
        <xsl:attribute name="alt"><xsl:value-of
          select="media-reference/@alternate-text"/></xsl:attribute>
      </xsl:element>
      <div align="right"><font size="-2">Photo:
        <xsl:value-of select="media-producer"/>
      </font></div>
      <b><font size="-1"><xsl:value-of select="media-caption"/></font></b>
    </td></tr>
  </xsl:element>
</xsl:template>

</xsl:stylesheet>
```

Appendix G

XML Applications

Area	Organization	XML Application
Accounting	American Institute of Certified Public Accountants (AICPA)	Extensible Financial Reporting Markup Language (XFRML)
Advertising	Newspaper Association of America (NAA)	NAA Classified Advertising Standards Task Force
Automotive	The Society of Automotive Engineers (SAE)	XML for the Automotive Industry - SAE J2008
Banking	Banking Industry Technology Secretariat (BITS)	Interactive Financial Exchange (IFX)
	Financial Services Technology Consortium (FSTC)	Bank Internet Payment System (BIPS)
	Microsoft, Intuit, CheckFree	Open Financial Exchange (OFX)
Communication	Alliance for Telecommunications Industry Solutions (ATIS)	Telecommunications Interchange Markup (TIM)
	Wireless Application Protocol Forum (WAP)	Wireless Markup Language (WML)
Content Syndication	Vignette, et al	The Information and Content Exchange Protocol (ICE)
Directory Services	The DSML Initiative	Directory Services Markup Language (DSML)
	Novell	DirXML
Distributed Management	Distributed Management Task Force, Inc.(DMTF)	Common Information Model (CIM)
Education	Educom IMS Project	IMS Meta-data Specification
	Schools Interoperability Framework	SIF
Electronic Commerce	CommerceNet	eCo Framework
	Commerce One	Common Business Library (CBL)
	CXML.org	Commerce XML (cXML)
	IBM	Business Rules Markup Language (BRML)
	Joint Electronic Commerce Program Office (JECPO)	Product Data Markup Language (PDML)
	MartSoft	Open Catalog Format (OCF)
	Open Trading Protocol group (OTP)	Open Trading Protocol (OTP)
EDI - Electronic Data Interchange	RosettaNet	RosettaNet
	Data Interchange Standards Association (DISA)	ANSI ASC X12/XML
	EEMA EDI/EC Work Group (CEN/ISSS)	XML/EDI Group
Enterprise Resource Planning	Open Applications Group (OAG)	Open Applications Group Interface Specification (OAGIS)
Financial	Financial Information eXchange protocol(FIX)	FIXML

Area	Organization	XML Application
Financial	FinXML.org	FinXML
	FpML.org	Financial Products Markup Language (FpML)
	Infinity	Network Trade Model (NTM)
Forms	JetForm Corporation	XML Forms Architecture (XFA)
	UWI.Com	Extensible Forms Description Language (XFDL)
Healthcare	Health Level Seven	HL7
	Phase Forward	Clinical Trial Data Model
Human Resources	DataMain	Human Resources Markup Language (hrml)
	HR-XML Consortium	JobPosting, CandidateProfile, Resume
	Open Applications Group (OAG)	Open Applications Group Interface Specification (OAGIS)
	Siemens Business Communication Systems	Siemens Time and Attendance System
	Tapestry.Net	JOB markup language (JOB)
Insurance	ACORD	Property and Casualty. Life (XMLife)
Intellectual Property Rights	Xerox Palo Alto Research Center	DPRL
Legal	U. S. District Court, District of New Mexico	XML Court Interface (XCI)
	Utah Electronic Law and Commerce Partnership	Legal XML Working Group
News	International Press Telecommunications Council (IPTC)	News Industry Text Format (NITF)
	XMLNews.org	XMLNews-Story, XMLNews-Meta
Real Estate	OpenMLS	Real Estate Listing Management System (OpenMLS)
	Real Estate Transaction Standard working group (RETS)	Real Estate Transaction Standard (RETS)
Science	NASA	Astronomical Instrument Markup Language (AIML)
	MoDL Project Team	Molecular Dynamics Markup Language(MoDL)
	The OpenMath Society	OpenMath
	Proteometrics	BIOPolymer Markup Language (BIOML)
	Peter Murray Rust	Chemical Markup Language (CML)
	Visual Genomics	Bioinformatic Sequence Markup Language (BSML)
	World Wide Web Consortium (W3C)	Mathematical Markup Language (MathML)
Software	IBM	Bean Markup Language (BML)
	INRIA	Koala Bean Markup Language (KBML)
	Marimba and Microsoft	Open Software Description Format (OSD)
	Object Management Group (OMG)	XML Metadata Interchange (XMI)
Travel	Open Travel Alliance	OTA

Area	Organization	XML Application
User Interface	Mozilla.org	Extensible User Interface Language (XUL)
	UIML.org	User Interface Markup Language (UIML)
Web Applications	Allaire	Cold Fusion Markup Language (CFML)
	Extensible Log Format project	Extensible Log Format (XLF)
	Internet Engineering Task Force (IETF)	Web Distributed Authoring and Versioning (WebDAV)
Workflow	Internet Engineering Task Force (IETF)	Simple Workflow Access Protocol (SWAP)
	Workflow Management Coalition (WfMC)	Wf-XML

