



NewsML 1 Requirements

Revision History

| | | |
|--------------------------|------------|--|
| 2nd IPTC Draft | 1999-10-01 | Created by Jo Rabin (Reuters). Presented to IPTC meeting at Amsterdam 1999-10-06, circulated under reference xn-2 . |
| 3rd IPTC Draft | 1999-12-10 | Updated following IPTC Standards Committee meeting at Heathrow 1999-11-22 |
| 4th IPTC Draft | 1999-12-21 | For Public Comment, edited by David Allen (IPTC) and Jo Rabin |
| 5th IPTC Draft | 2000-03-16 | Updated following Washington meeting 2000-01-28, circulated under reference nsm0002 . |
| 1st revision | 2000-03-22 | Updated in Nice, circulated under reference nsm0007 . |
| 2 nd revision | 2004-04-08 | Edited by Laurent Le Meur (AFP). The original term "part" replaced by the term "component" for better consistency with final NewsML definitions. Minor re-numbering and presentation update. |

Headlines

| Code | Description |
|------|--|
| 100 | Support the representation of electronic news entities i.e. <i>news-items</i> , components of news-items, collections of news-items, relationships between news-items and metadata associated with news-items. |
| 200 | Be usable throughout the news life cycle. |
| 300 | Allow news-items to consist of arbitrary mixtures of media types, languages and encoding. |
| 400 | Be usable either as a replacement for or allow the transport of all existing news formats and encodings. |
| 500 | Support a number of different constructions of the same data |
| 600 | Support the management of news-items and their development over time |
| 700 | Be simply extensible and flexible. |
| 800 | Allow for authentication and signature of metadata and news-item content. |
| 900 | Not be unduly verbose. |
| 1000 | Use XML and other appropriate standards and recommendations. |
| 1100 | Provide a lightweight facility for the representation of text |

Definitions

- News-item** A unit of news as defined by the editorial or product construction practice of a supplier.
- Component** A part of a news-item, which may consist of other news-items, pieces of text, pictures, etc. A component bears a specific named relationship to its container. Components may contain alternatives, which fulfill that named role, or complements which bear subsidiary named relationships to the container.

Development and Explanation of Requirements

Note: the current numbering is the result of many modifications of this requirement expression.

100 **Support the representation of electronic news entities i.e. news-items, components of news-items, collections of news-items, relationships between news-items and metadata associated with news-items**

- 110 **Structure:** News-items have structure, i.e. are a composite of components of arbitrary media types with named relationships to each other.
- 115 **Links:** News-items may have named relationships to other news-items.
- 116 **Named Entity Identification:** The presence of and the location of named entities (i.e. people, places, products etc) within news-items must be supported in a media independent manner.
- 120 **Specific Features of News:** News-items of all media types have specific features like by-line, dateline, slug ...
- 150 **Metadata:** NewsML must be capable of representing metadata necessary to support news as it passes through its lifecycle (see note).
- 151 **Metadata Standards:** NewsML must be capable of representing metadata from both standard and non-standard schemes as well as alternative representations of the same metadata (e.g. organized according to different schemes).
- 152 **External Metadata:** NewsML must support the attachment of metadata that is relevant only to a specific application or production process. This metadata need not be XML formatted.
- 160 **Electronic News:** News ML does not natively support features required specifically for news production in a particular medium (e.g. print) but does support the reference to this information as an external scheme (cf. 152).
- 170 **Presentation Semantics:** NewsML does not natively convey presentation semantics but supports the use of standard mechanisms that do support such semantics, such as style sheets. It covers logical components but does not specify the physical or temporal relationship between components when presented in a publishing medium.
- ### 200 **Be usable throughout the news life cycle.**
- 210 **Syndication:** NewsML must be capable of representing the semantics required for news which pass through an arbitrarily complex chain consisting of original publisher and any number of other integrators, aggregators and distributors. Value may be added at any point in this distribution chain.
- 230 **Construction Tools:** NewsML must be suitable for use with off-the-shelf editors and other applications for manipulation of news without undue configuration and with minimal customisation of those applications.

- 240 **Display Tools:** NewsML must be deliverable to commonly available consumer applications, especially browsers, without undue configuration and with minimal customisation of those applications.
- 250 **Transformation:** It must be possible to transform NewsML into and from a range of formats, especially IIM, and NITF. Since NewsML provides for considerably greater richness of structure than other formats such transformations may involve a loss of information. NewsML must also be transformable into delivery formats such as HTML and WML etc.
- 260 **Workflow:** NewsML does not natively support workflow semantics. It does support the use of standard mechanisms that do support such semantics to allow routing of news-items through the editorial and production processes. NewsML allows for the capture and retention of some aspects of workflow process information as "production metadata".
- 300 Allow news-items to consist of arbitrary mixtures of media types, languages and encodings**
- 310 **No Text Bias:** NewsML is not to assume that text is the primary vehicle for news. All media types are to be treated equally.
- 320 **Alternatives:** NewsML must support alternative renderings of the same content as equivalents - e.g. text in different languages, text in different encodings graphics in different encodings. Encoding refers to the application format of the data (e.g. HTML or MSWord) as well as what type of compression, encryption or technique for representation of the value of the data (base64, binary or whatever) It must be possible to determine the original representation of a content.
- 325 **Source Language:** Where more than one language alternative is available it must be possible to identify the original alternative from which a translation is derived.
- 330 **Heterogeneous Parts:** When considering relationships between the components of News-items, no assumptions are made as to the media types of those components, nor is it to be assumed that alternative renderings of the same component are of the same media type.
- 340 **International:** All languages that can be represented using Unicode are supported equally.
- 400 Be usable either as a replacement for or allow the transport of all existing news formats and encodings**
- 410 **Replacement:** NewsML must be able to act as a replacement for existing formats (e.g. IIM) where these formats fall within the requirements of NewsML.
- 420 **Transport:** Where existing formats (including IIM) convey semantics beyond those of NewsML such as layout, NewsML must be able to transport objects in such formats as alternative representations.
- 430 **Transmission Capability:** NewsML is required to be suitable for streaming and broadcast (i.e. essentially one-way) environments as well as request/response environments of varying forms.
- 500 support a number of different constructions of the same data**
- 510 **Heterogeneous Representation:** The constraints of time and space as well as the demands of specific applications and delivery environments make it necessary to support different physical representations of the same news-item. Some applications may demand the delivery of a News-item and its components as a unitary entity (e.g. in a streaming or broadcast environment), whereas others require that only the minimum information be transferred and that additional information in the news-item be retrievable on demand (e.g. in a mobile WAP environment).

- 520 **Inclusion/Exclusion:** To provide this flexibility NewsML must support the following concepts:
Explicit Inclusion: Information whose values are encoded in a particular representation of a news-item.
Inclusion by Reference: Information about how to obtain the values of the information in question.
Exclusion: Systematic or temporary omission of a component or class of data in a news-item as a feature of an application. (e.g. exclusion of certain types of metadata when publishing, exclusion of categorization and text information when retrieving lists of headlines ...).
- 525 **Mandatory Features:** NewsML will specify a minimal number of mandatory features. The absence of a non-mandatory feature does not imply that the feature does not exist.
- 530 **Identification:** Because the component parts of a news-item may be included and excluded in arbitrary combinations NewsML must provide an identification mechanism through which it can be determined that two news-items are the same or not, without referring to their content. The identification is to be unambiguous and unique. The decision on assignment of identity is an editorial decision.
- 540 **Nomenclature:** To assist with the discussion of varying forms of representing news the following terms are used with specific meanings:
Representation: Two entities with corresponding parts included and excluded have the same representation. Representation refers to the physical form something has.
Revision: Two entities that have the same identifier and the same revision label are at the same revision. Revision refers to how up to date something is and does not refer to the physical form of presentation.
- 600 support the management of news-items and their development over time**
- 610 **Takes:** NewsML must support the creation and incremental delivery of partial news-items as well as the consolidated delivery of partial news-items.
- 620 **News-item Development:** It must provide for the changing construction of news-items over time, which may result from the different timeliness of availability of different media types (e.g. text may precede audio may precede video).
- 630 **Maintenance:** NewsML must support news maintenance tasks such as deletions, corrections, overwrites, embargoes and so on.
- 640 **Revisions:** NewsML must provide a deterministic mechanism to allow applications to distinguish earlier revisions of news-items from later revisions.
- 700 Be extensible and flexible.**
- 710 **Evolutionary Approach:** The requirements of NewsML will develop over time. It must be possible to develop NewsML to meet those requirements without major discontinuity in the applications that support it.
- 720 **Include Non News Information:** Provide for inclusion of advertisements, sponsorship notices and application specific information - for example it must be possible to include processable information from non-news domains (such as stock quotes) for applications that need such things. Presence of this data must not present an obstacle to rendering or other processing of NewsML by applications that were not designed with such features in mind. Default behaviour should be established to allow sensible rendering in the absence of an application specific processor.
- 730 **Engineering Principles:** To facilitate coherent and consistent development of NewsML it must have stated design principles (which must include naming conventions, see 1020).

740 **Documentation:** As far as is practical the rationale behind design decisions made during the course of its development must be available in the future. Changes should be identified through a suitable audit trail.

800 **Allow for signature and authentication of metadata and news-item content.**

810 **Signed Content:** NewsML must support the digital signature and authentication of content. Different parts of the same news-item may have different signatures.

820 **Signed Metadata:** NewsML must support the digital signature and authentication of metadata. Since metadata is applied at various points in the development of a news-item multiple signatures must be supported.

900 **Not be unduly verbose.**

910 **Transmission Capacity:** NewsML is to be used in a wide variety of contexts with varying transmission capabilities. NewsML needs to provide the ability to express semantics in as concise a way as is possible while retaining the required flexibility. This is particularly important in respect of the development of news-items (above). It should be possible, at a later time, to deterministically recover such information in a comprehensive manner at a news-item level.

920 **Defaults:** Provide for the setting of defaults to minimise duplication and repetition of data.

930 **Special Case for Text:** Since most existing news-items have extremely simple structure (they have a single textual component), the overhead for representing extremely small news-items should be reduced as far as possible.

940 **Binary Data:** NewsML must support the inclusion of binary parts in their native form (e.g. as well as supporting a Base 64 encoded representations of this data native binary representations should also be possible).

950 **Ease of Processing:** So far as is possible the encoding should be designed with ease and efficiency of processing by receivers in mind.

1000 **Use XML and other appropriate standards and recommendations.**

1001 **Web Conformance:** Be consistent with widely adopted techniques and published international standards and recommendations to gain maximum leverage from the intellectual effort that has gone into their creation. In addition to gain maximum leverage from utilities, components and tools that are widely available for their processing.

1010 **Evolution:** Incorporate new standards as they become available.

1020 **Use of XML:** A consistent naming convention is to be adopted for the naming of elements and attributes. The naming conventions must not be unduly verbose, while maintaining readability.

10646 **Universality:** Use Unicode where practical.

1100 **Provide a lightweight facility for the representation of text**

1105 **Content of news-lines:** These "specific features of news" referred to in 120 are inherently textual and need to be represented in an expeditious manner.

1110 **Text Structure:** Provide a means of identifying the structure and sub-structure within textual parts (paragraphs, tables etc.).

1120 **Text Markup:** The determination of the scope of this facility will be through the application of the following tests:

1 Is there a parallel structure in NITF that can be used?

2 Can this mechanism be applied equally to all media?

3 Do the structures fall in with the overall NewsML design requirements especially of minimising the number of tags?

Notes

Collections

- N110 News-items are very frequently found together in collections. NewsML supports collections of news objects that have been assembled with journalistic intent (such as a composite news-item consisting of the top ten news-items of the hour) and collections of news objects that are constructed in a more arbitrary way (such as responses to queries).
- N120 Collections are represented as news-items.
- N130 The intention behind the construction of collections may be determined from the metadata.

Equivalence of News-items

- N210 Because of the equivalence of alternatives and the ability to exclude data from any instance of a news-item, the equivalence or otherwise of news-items cannot be established from their physical form. Hence also there is no standard representation [or canonical form] for a news-item.
- N220 NewsML provides a mechanism for providers to mark news-items so that they may be distinguished in a unique and unambiguous way without reference to their representation.

Permissioning

- N310 Permissioning is not considered to be part of news management. It will be applied through external applications.

Life Cycle

- N410 Reference is made to the notion of a news life cycle. This notion is intended to capture the characteristic processes through which news-items may pass between being created and being consumed. This life cycle includes but is not restricted to assignment, authoring, editing, storage, archiving, distribution, being searchable and so on.
- N420 NewsML does not provide the ability to specify or support workflow. It does provide the ability to capture the creation and distribution history of a news-item.