

```
<NewsML>  
<Catalog Href=" ../catalog-nsk/catalog.NskMasterCatalog.xml" />  
<NewsEnvelope>  
  <TransmissionId>adc123</TransmissionId>  
  <SentFrom>  
    <Party FormalName="NskSportsSection" />  
  </SentFrom>
```

# News用XMLの世界標準 - NewsML -



## メディア業界におけるXMLの取り組みと規格内容

XMLコンソーシアム・ドキュメント部会

2001年 11月 22日 朝日新聞社映像センター 竹原 大祐

日本アイ・ビー・エム 藤原 隆弘

(日本新聞協会 新データフォーマット策定チームメンバー)



# はじめに

- XMLコンソーシアム・ドキュメント部会ではドキュメント関係のXMLに興味のある人たちが構成されています。
- この資料は、今までにドキュメント部会で紹介された中で関心度の高かった一つ、「NewsML(ニューズエムエル)」を、NewsMLに詳しいドキュメント部会メンバーが中心に中間発表会向けにまとめたものです。

# NewsML誕生の背景



# メディア環境の変化 コンテンツの多様化

- ニュースソースの多様化
- コンテンツの多様化
  - AP (広告、放送)、ロイター (金融サービス)
  - CNN (ビデオ、WEB)、ブルームバーグ (経済)



**Associated Press**

AP Broadcast  
AP FINANCIAL NEWS SERVICES  
AP PHOTO ARCHIVE  
AdSEND

**REUTERS** 

金融プロダクト・ロイター3000シリーズ  
ロイター・ファイナンシャル・ビデオ・オンライン  
オンライン・マーケット・ニュース  
ピクチャー・アーカイブ・オンライン  
ワールド・スポーツ・オンライン・レポート

**CNN.com**

**Bloomberg**  
NEWS

# メディア環境の変化 媒体の多様化と標準化

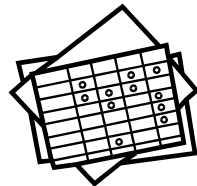
## ■ 配布技術の多様化

- 専用網 → 無線、衛星、携帯、メール、WEB



## ■ フォーマットの多様化

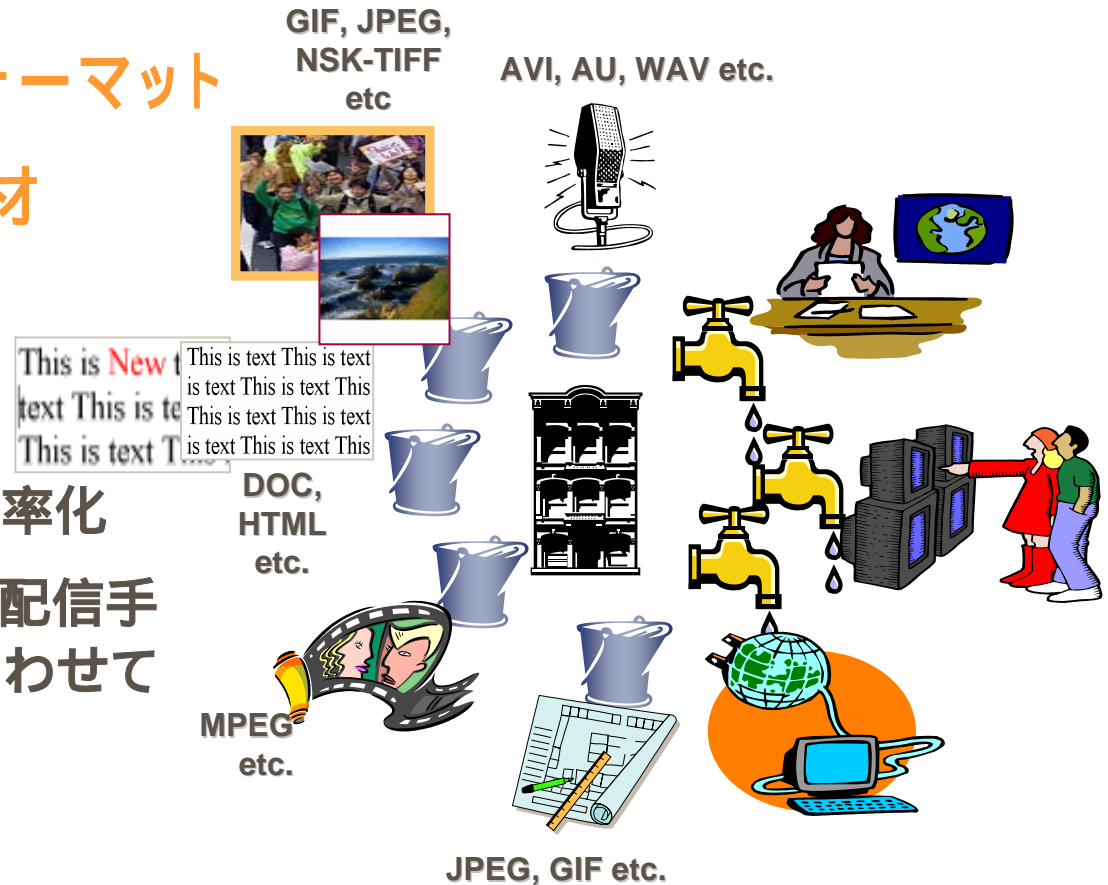
- ベタ打ち、バイナリー → CSV、タグ付き → HTML、XML



# メディア環境の変化 コンテンツのフローのマルチ化

- 収集 あらゆる**フォーマット**
- 統合 あらゆる**素材**
- 更新 **最新の版**

1. 素材の加工フローを効率化
2. 素材のパッケージ化 配信手段や配信チャンネルにあわせて



# メディア環境の変化 コンテンツ・マネージメント



- 集配信 直線から同心円に
- 素材加工 同心円の中心で新しいソリューション  
DAM デジタル・アセット・マネージメント
- 組版 素材のリンクで編集の効率化

# 新聞社・通信社のニーズ

- ニュースを集める、加工する
  - 多様なフォーマットの素材
  - 写真や記事の著作権
  - 素材を集約・編集・検索・保存
  - 時系列でニュースを追いかける
  - 簡単で安価な素材流通システム
- 次々と生まれてくるフォーマットとニーズに柔軟に対応  
ニュースのフォーマット統一が重要



# NewsMLの誕生

- 98年、ロイターのマーク・ハント氏、ジョウ・レイベン氏が定義
- 99年10月 IPTCがNewsMLを標準化

**IPTC:世界約55社の通信社・新聞社・ベンダーが標準化と開発に参加**(ニューヨークタイムズ、AP、共同通信など)

(International Press Telecommunication Council) **国際新聞通信評議会**

- 00年初 IPTCが機能定義、DTD作成
- 00年10月 IPTCでバージョン1.0発行
  - マルチメディア・コンテンツの新しい市場へ



# 日本新聞協会

- 新聞倫理綱領を制定し実践する自主組織  
1946年7月に創立した社団法人
- 新聞112、通信4、放送39(ラジオ単営8、テレビ単営25、ラ・テ兼営6)計155社
- 各種フォーマット策定のためベンダー参加
- 各社の代表者で構成する総会、理事会のもとに、各種の委員会、専門部会が設置

# 誕生 日本新聞協会

- 00年6月 NewsMLの分析と解説
  - 「新データフォーマット策定チーム」
  - NewsMLに沿った共通認識
  - NewsMLの仕様書の翻訳
  - 日本固有の属性情報の定義
- 01年8月  
**NSK-NewsMLレベル1公開**



# XMLベースの特徴

- アプリケーションに依存しない
- ファイルフォーマットが統一  
速報性や扱いが容易に
- XMLとしての拡張性
- データの構造化が容易
  - 封筒や送受信情報の取り込み

## XMLベースの特徴(2)

- データの妥当性チェックが容易
    - スキーマ型の検討も
  - テキストベースであるメリット
    - ブラウジング
    - トラブル対応
    - アプリケーション処理
- ユーザースキルをサポート



# NewsMLの特徴

- ボキャブラリ = システム間の用語定義
- カタログ = ボキャブラリの参照方法
- 構造 = ニュース素材を扱う
- 時系列 = 刻々と更新するニュース



## NewsMLの特徴(2)

- 複数のメディアを同時に扱える
- どんなメディアでも公平に扱う
- 物理データは複数扱う エンティティ参照、ポインタ機構
- 素材の発生元がわかる
- 著作権情報、記事の再利用指示

## NewsMLの特徴(3)

- ユニコード対応 = 日本語も可
- 関連記事の参照が可能
- 標準化団体 = IPTC
- ボキャブラリの統一
  - カテゴリによる検索が可能
- 履歴管理で速報性の強化



# NSK-NewsMLレベル1の制限

- NewsML1.0DTDとIPTCの仕様に準じる
- 素材部分には
  - 日本語の要素名は禁止
  - NewsMLと同じ要素名は禁止
  - DTDを規定する場合には外部参照
- Validな文書
- UPDATEは使わない

# NewsMLの使用例



# 欧米の応用例



- IPTCサイトでDTD、機能仕様を公開
- 日本新聞協会には日本語版の仕様書
- ロイターでWEBのデモを公開
  - 最新のニュースをNewsML化
- ロイター:IDS (Internet Delivery System) の利用
  - 02年2月から データをPULL型でダウンロード
- 他の通信社も採用 BusinessWire、Press Association、WSJ.com



# 国内の応用例

- 新聞各社 新幹線の電光掲示板



- 毎日 WEBサイトでデモ。情報公開も

- 共同 03年に配信をNewsMLで



# 今後のNewsML



# XMLコンソーシアム・応用部会

## ■ NewsMLワーキンググループ

リーダー：朝日新聞・竹原

サブリーダー：IBM・藤原

- XMLベースでの研究・問題点の洗い出し
- ツールキットの開発
- アプリケーション構築と検証
  - データ構築、閲覧、データベース
  - クライアントサイド・アプリケーション

# NewsMLの今後の方向性

- IPTC 仕様策定
- 日本新聞協会 IPTCへ要求をアップする
- DTDでのデータ妥当性検証
  - 情報の正確性、厳密性の担保ができるか
- スキーマへの対応
- 他のXMLとの連携
- セキュリティ・デジタル署名
- XMLの技術とともに進化する

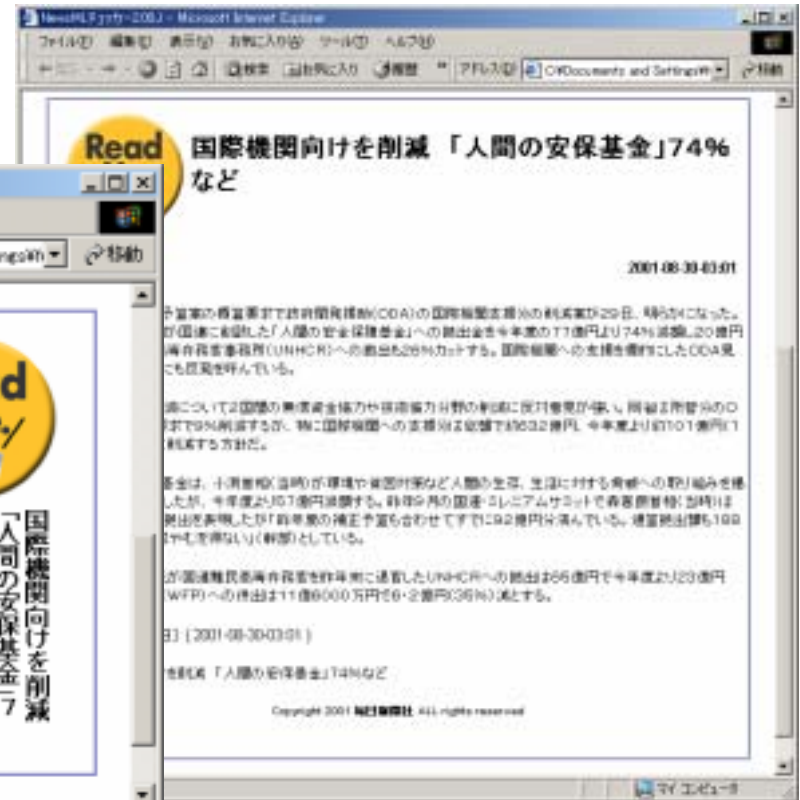
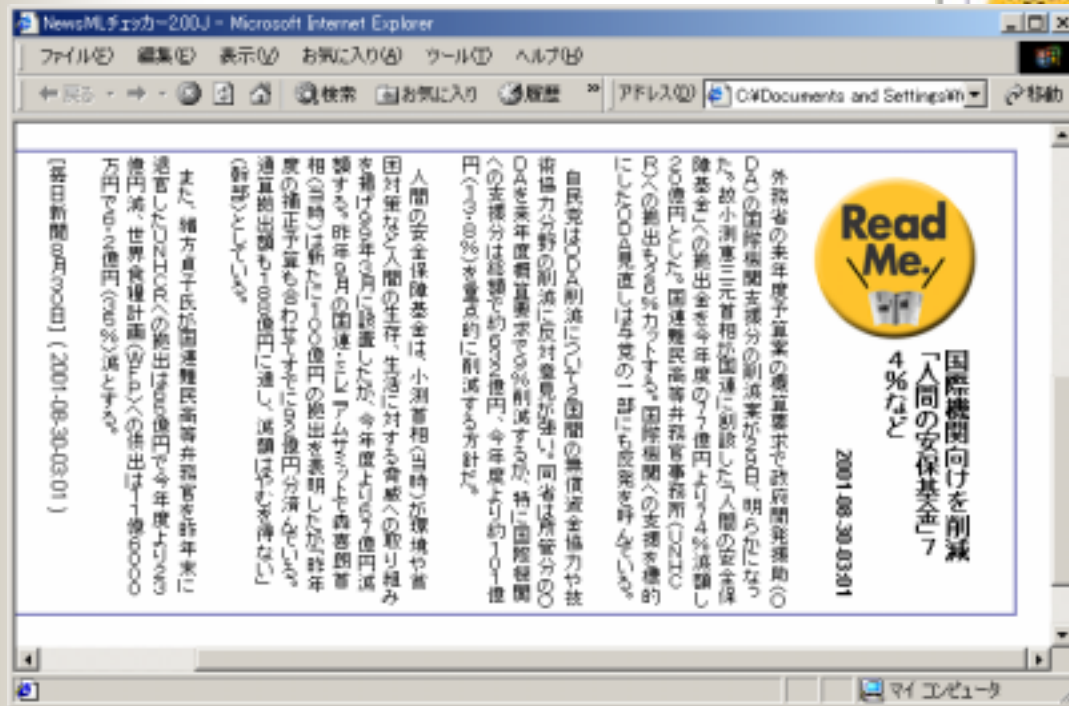
# NewsMLの構造

```
<NewsItem>
  <Identification>
    <NewsIdentifier>
      <ProviderId>pressnet.or.jp</ProviderId>
      <DateId>20010602</DateId>
      <NewsItemId>sportsnews123456</NewsItemId>
      <RevisionId PreviousRevision="1" Update="N">2</RevisionId>
      <PublicIdentifier>urn:newsml:pressnet.or.jp:20010602:sportsnews1234
    </NewsIdentifier>
  </Identification>
  <NewsManagement>
    <NewsItemType FormalName="News" />
    <FirstCreated>20010602T200012+0900</FirstCreated>
    <ThisRevisionCreated>20010602T220043+0900</ThisRevisionCreated>
    <Status FormalName="Usable" />
    <DerivedFrom NewsItem="urn:newsml:pressnet.or.jp:20010601:sportsnews123456">
      <Comment xml:lang="ja">前日のsportsnews012345</Comment>
    </DerivedFrom>
    <AssociatedWith NewsItem="urn:newsml:pressnet.or.jp:20010602:photo20010602">
      <Comment xml:lang="ja">併用写真</Comment>
    </AssociatedWith>
  </NewsManagement>
  <NewsComponent Duid="NC0001" xml:lang="ja">
    <Comment>関連写真があります</Comment>
    <NewsLines>
      <HeadLine>鈴木2得点</HeadLine>
      <SubHeadLine>日本準決勝進出</SubHeadLine>
      <ByLine>新聞太郎</ByLine>
      <DateLine>2001年6月2日、新潟県、新潟スタジアム</DateLine>
      <CreditLine>Pressnet</CreditLine>
      <CopyrightLine xml:lang="en">NSK all rights reserved.</CopyrightLine>
      <RightsLine>2次使用禁止、新聞紙面用メディア使用禁止</RightsLine>
      <SeriesLine>コンフェデ杯特集</SeriesLine>
      <KeywordLine>コンフェデ杯</KeywordLine>
      <KeywordLine>サッカー</KeywordLine>
    </NewsLines>
    <NewsLine>
      <NewsLineType FormalName="Karimidashi" />
      <NewsLineText> サンプル</NewsLineText>
    </NewsLine>
  </NewsComponent>
  <RightsMetadata>
    <Copyright>
      <CopyrightHolder>日本新聞協会</CopyrightHolder>
      <CopyrightDate>2001-2002</CopyrightDate>
    </Copyright>
  </RightsMetadata>
  <DescriptiveMetadata>
    <Language FormalName="ja" />
    <SubjectCode>
```



# NewsML表示サンプル

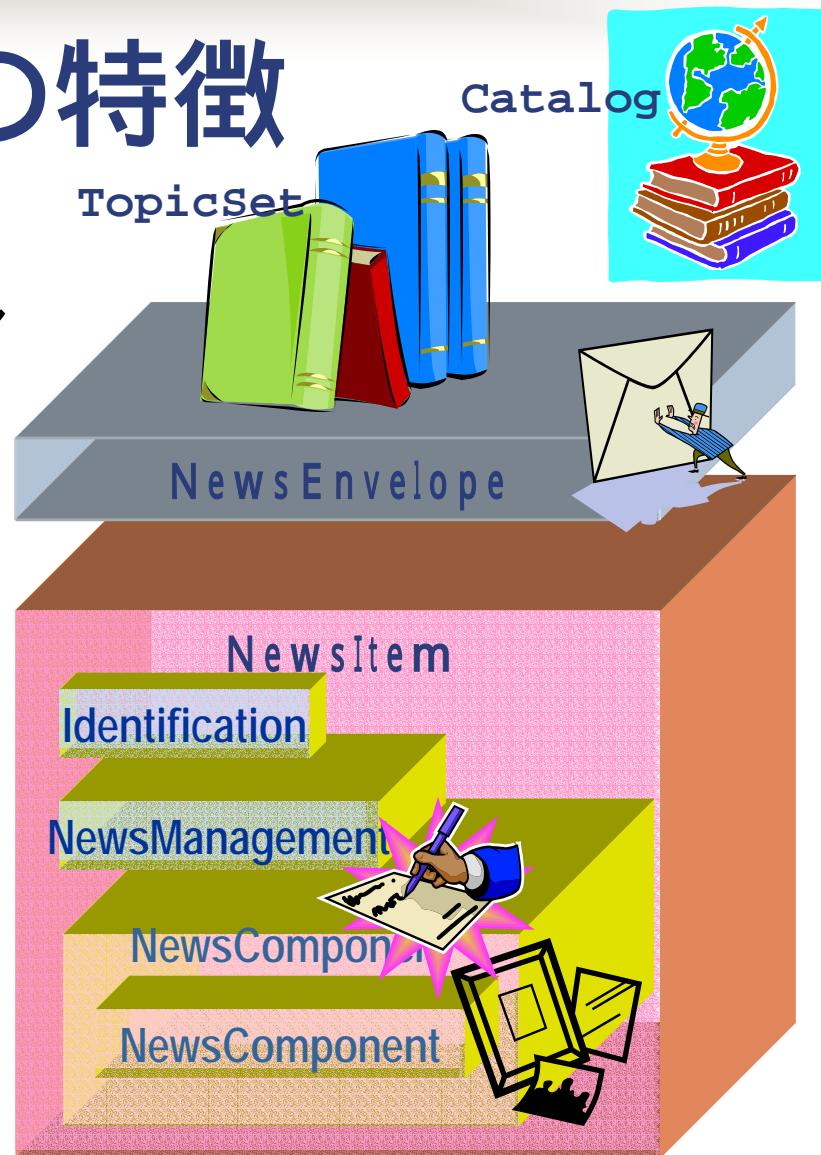
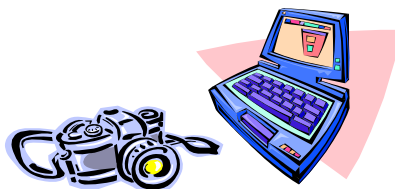
- 同じ内容を横書き・縦書き・携帯で...



# NewsML構造の特徴

## ■ データ構造

- NewsEnvelope 封筒のラベル
- NewsItem ニュース管理単位
  - Identification ID情報
  - NewsManagement 管理情報
  - NewsComponent ニュース構成要素
    - ContentItem 素材構成要素
    - DataContent 実データ
- TopicSet 辞書
- Catalog 辞書の目録



# 実際のNewsMLファイル(1)

## ■ [NewsEnvelope (封筒) まで]

<NewsML>

<Catalog>

<Resource>

<Urn>urn:newsml:pressnet.or.jp:20010516:topicset.iptc-format-nsk:1</Urn>

<Url>../topicsets/topicset.iptc-format-nsk.xml</Url>

<DefaultVocabularyFor Scheme="IptcFormats"

Context="Format/@FormalName" />

</Resource>

...

</Catalog>

<NewsEnvelope>

<Transmissionid>photo123</Transmissionid>

<SentFrom>

<Party FormalName="NSK" />

</SentFrom>

<SentTo>

<Party FormalName="GUEST" />

</SentTo>

<DateAndTime>20010602T230000+0900</DateAndTime>

<NewsService FormalName="SPORTSPHOTO" />

<NewsProduct FormalName="NskNewsML:1" />

<Priority FormalName="5" />

</NewsEnvelope>

<NewsItem>

...

辞書の目録

送信ID

送信元

宛先

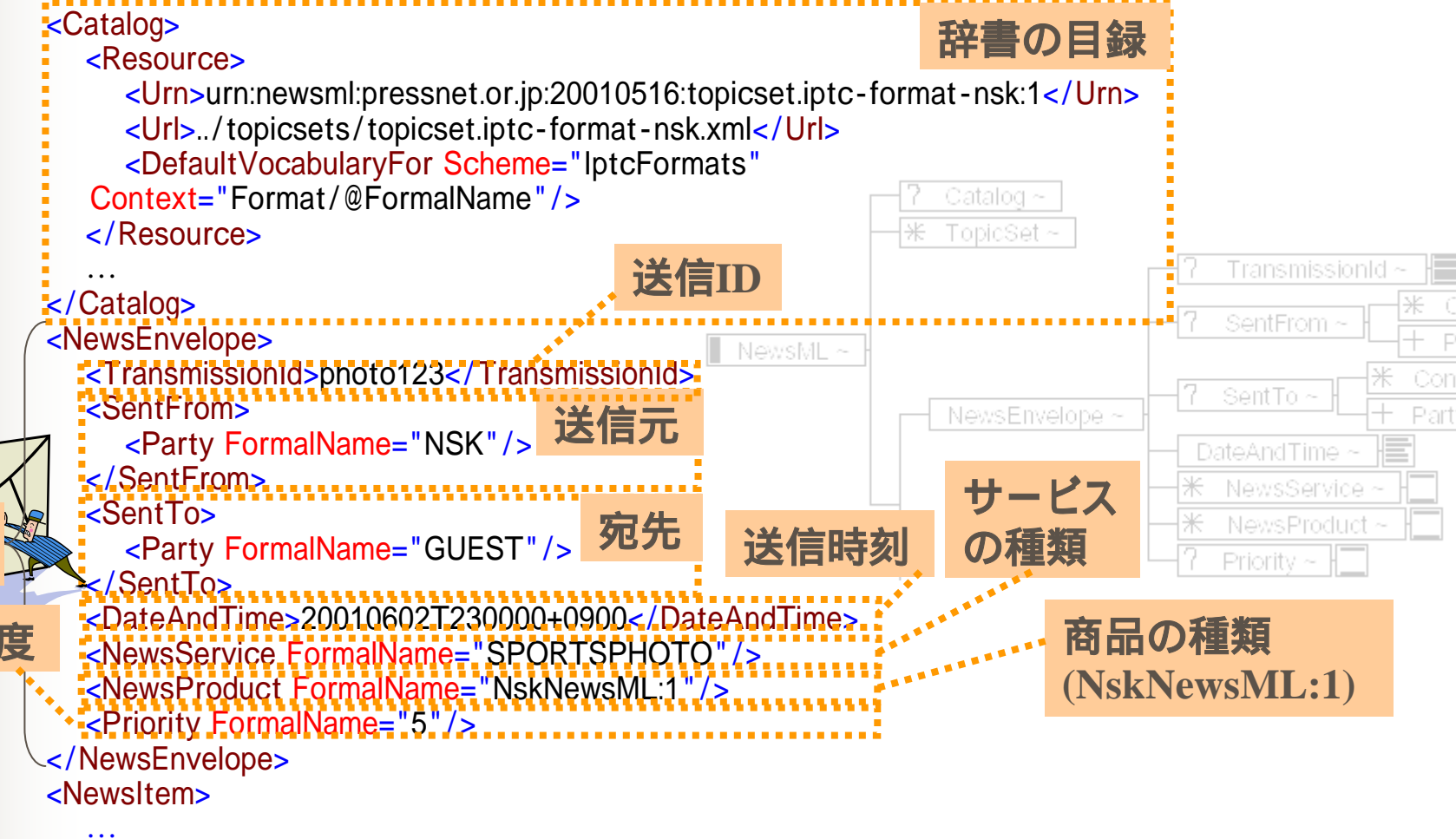
送信時刻

サービスの種類

商品の種類  
(NskNewsML:1)

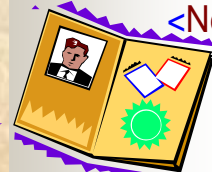
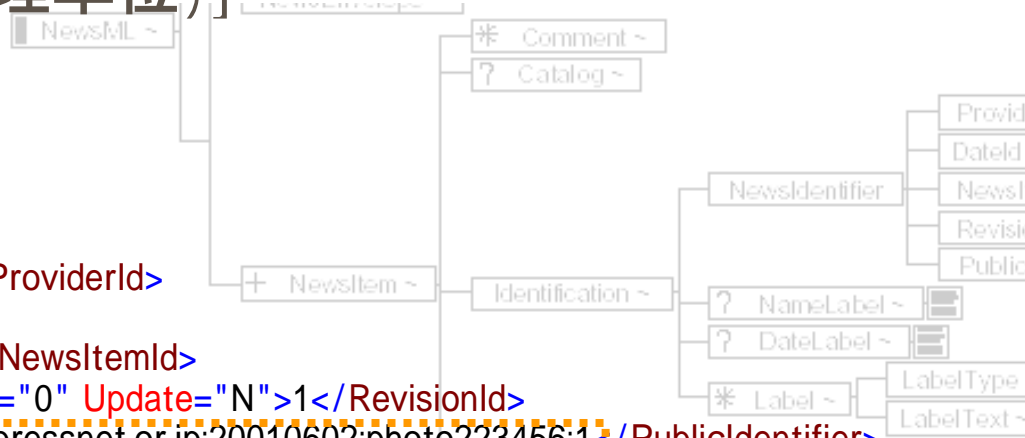
封筒のラベル

優先度



# 実際のNewsMLファイル(2)

## ■ [NewsItem (ニュース管理単位)]



ID情報

```

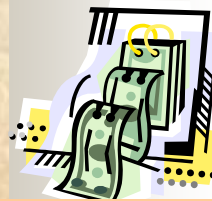
<NewsItem xml:lang="ja">
  <Identification>
    <NewsIdentifier>
      <ProviderId>pressnet.or.jp</ProviderId>
      <Dateld>20010602</Dateld>
      <NewsItemId>photo223456</NewsItemId>
      <RevisionId PreviousRevision="0" Update="N">1</RevisionId>
      <PublicIdentifier>urn:newsml:pressnet.or.jp:20010602:photo223456:1</PublicIdentifier>
    </NewsIdentifier>
  </Identification>
  <NewsManagement>
    <NewsItemType FormalName="News" />
    <FirstCreated>20010602T200024+0900</FirstCreated>
    <ThisRevisionCreated>20010602T200024+0900</ThisRevisionCreated>
    <Status FormalName="Usable" />
    <AssociatedWith NewsItem="urn:newsml:pressnet.or.jp:20010602:sportsnews123456">
      <Comment xml:lang="ja">関連記事</Comment>
    </AssociatedWith>
  </NewsManagement>
  <NewsComponent >
  ...
  
```

NewsItemの種類

NewsML URN (後

版の利用可能状態 (後述)

リンク情報 (後述)



管理情報

ニュース構成要素



# NewsComponent

NewsComponent



## ■ ニュース構成要素

- AdministrativeMetadata  
ニュース構成 管理情報
- DescriptiveMetadata  
説明情報
- RightsMetadata  
権利情報
- Metadata  
汎用付加情報
- NewsLines  
編集者情報
- ContentItem  
素材構成要素
  
- NewsComponent (子)
- NewsItem (子)



# NewsMLの構造

## NewsComponentとContentItem

NewsComponent



- 配信元,
- 作成元,
- 情報元,
- 加工情報,
- 履歴

ニュース構成管理情報

- 言語, ジャンル,
- 背景, コメント,
- 関連情報,
- 世代情報

説明情報

- 権利所有者
- 権利情報

権利条文

- 見出し 著者 キーワード その他見出し
- 編集者情報

素材データ(実データ)or  
(ファイル・ポインター)

- サイズ, 動画の時間
  - 幅, 高さ, 解像度
  - 色, ストリーミング
- 物理 メタデータ

- メディアタイプ
- 素材データ情報

ContentItem

NewsComponent

# 実際のNewsMLファイル(3)

## NewsComponent[ニュース構成要素]

NewsComponent



```
<NewsComponent Duid="NC001" xml:lang="ja">
```

```
<NewsLines>
```

```
<HeadLine>歓喜する鈴木</HeadLine>
```

```
<SubHeadLine>日本準決勝進出</SubHeadLine>
```

```
<ByLine>写真太郎</ByLine>
```

```
<DateLine>2001年6月2日、新潟県、新潟スタジアム</DateLine>
```

```
<NewsLine>
```

```
<NewsLineType FormalName="Caption" />
```

```
<NewsLineText>カメルーン 日本 前半8分、先制ゴールに歓喜する鈴木</NewsLineText>
```

```
</NewsLine>
```

```
</NewsLines>
```

```
<RightsMetadata>
```

```
<Copyright>
```

```
<CopyrightHolder>日本新聞協会</CopyrightHolder>
```

```
<CopyrightDate>2001-2002</CopyrightDate>
```

```
</Copyright>
```

```
</RightsMetadata>
```

```
<DescriptiveMetadata>
```

```
<SubjectCode>
```

```
<Subject FormalName="15000000" />
```

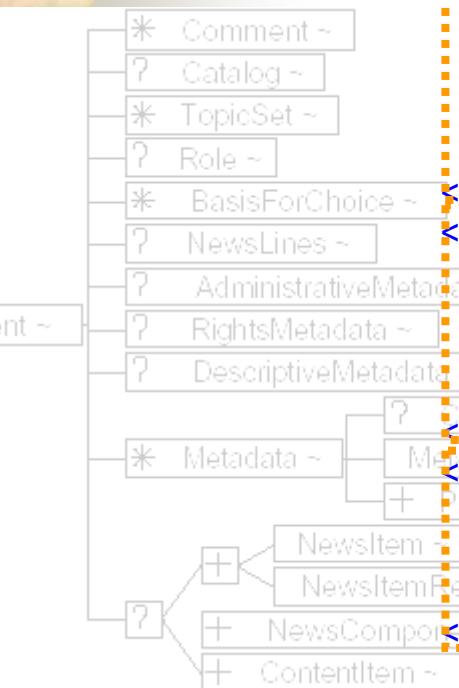
```
</SubjectCode>
```

```
</DescriptiveMetadata>
```

編集者情報

権利情報

説明情報



# 実際のNewsMLファイル(4)

## ContentItem[素材構成要素]



<NewsComponent>

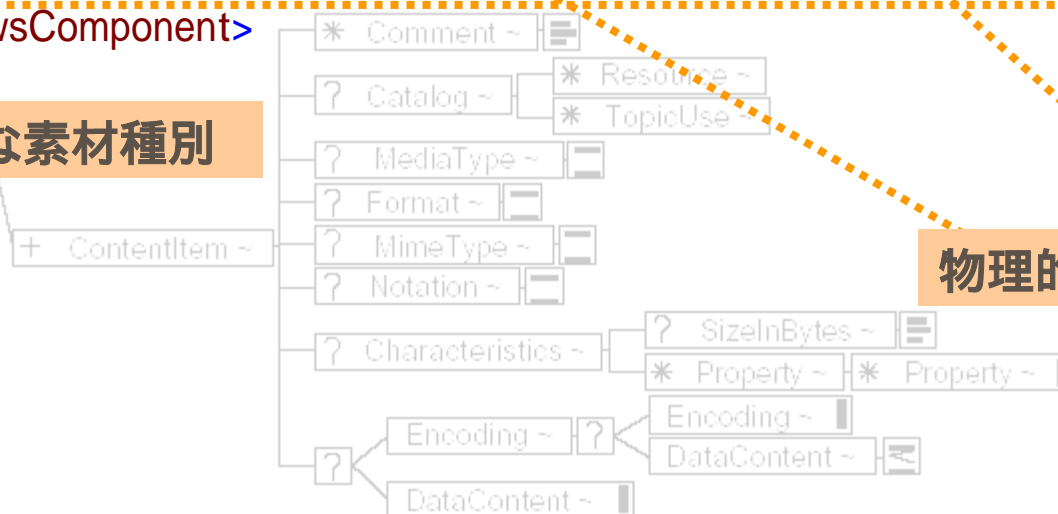
```
...
<ContentItem Href="./Sample.jpg">
  <MediaType FormalName="Photo"/>
  <MimeType FormalName="image/jpeg"/>
  <Characteristics>
    <SizeInBytes>15232</SizeInBytes>
  </Characteristics>
</ContentItem>
</NewsComponent>
```

素材構成要素

感覚的な素材種別

MimeType  
やFormat

物理的な属性情報





# NewsMLの構造

## Topic Set



### ■ ボキャブラリ集

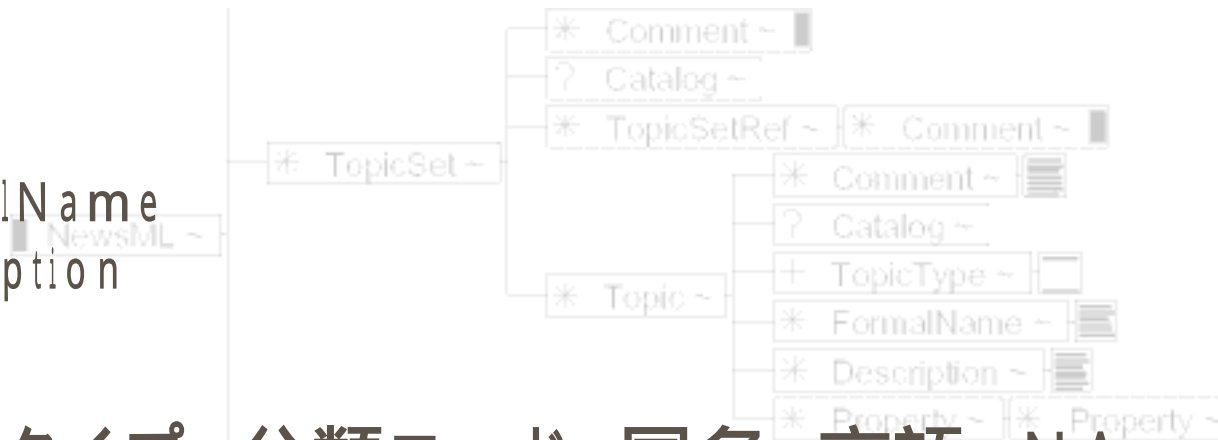
- ニュースの付加情報として利用

### ■ 構造

- Topic
  - FormalName
  - Description

### ■ 種類

- メディアタイプ、分類コード、国名、言語、NA SDAQ企業コードなど



# NewsMLの構造

## Topic Set (2)



信頼度  
素材重要性  
利用可能状態

緊急度  
メタデータの重要性  
サブジェクトコード(大分類、中分類)

配信優先順位  
読者との関連性

フォーマット  
MimeType  
メタデータ

記法  
メディアタイプ  
ニュースアイテムタイプ

ニュースラインタイプ  
トピックタイプ  
ロール

プロパティ  
ジャンル  
言語

配信団体  
NewsItemのラベル  
国名

年齢層  
プロダクト  
通貨

団体(NSKTiff登録)  
米国産業コード

日本の地域  
ナスダック会社コード

# Topic Setの例と指定方法

Status: - 利用可能状態 -

## ■ 種類

- Usable (公開可能)
- Embargoed (公開待機)
- Canceled (公開取消)
- Withheld (公開未定)

## ■ 指定方法

```
<NewsManagement>  
  <Status FormalName="Usable" />
```

## ■ 定義 (TopicSetファイル)

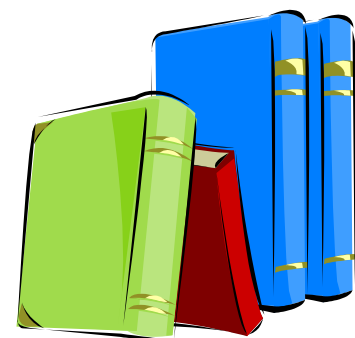
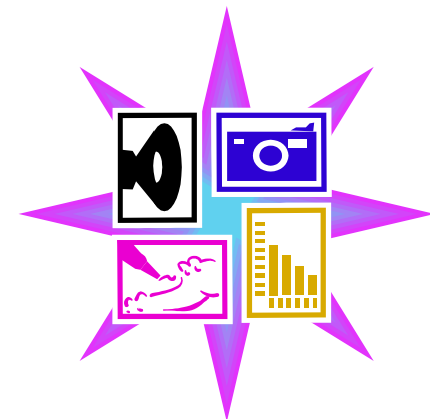
```
...  
</NewsManagement>  
<TopicSet Duid="iptc.status" FormalName="Status">  
  <Comment xml:lang="en">The current usability of a NewsItem</Comment>  
  <TopicSetRef TopicSet="urn:newsml:iptc.org:20001006:topicset1">  
    <Topic Duid="stat1">  
      <TopicType Scheme="IptcTopicType" FormalName="Status">  
        <FormalName Scheme="IptcStatus">Usable</FormalName>  
        <Description xml:lang="en">The NewsItem and its content are available for  
        <Description xml:lang="ja">NewsItem及びその内容物に  
      </Topic>  
    <Topic Duid="stat2">  
      <TopicType Scheme="IptcTopicType" FormalName="Status">  
        <FormalName Scheme="IptcStatus">Embargoed</FormalName>  
        <Description xml:lang="en">Neither the NewsItem nor its content are available for  
        publication by the provider.</Description>  
        <Description xml:lang="ja">NewsItem及びその内容物に  
      </Description>  
    </Topic>  
  </TopicSetRef>  
  ...  
</TopicSet>
```

# NewsMLの重要機能



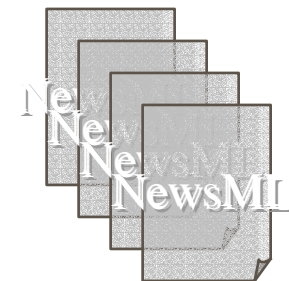
# NewsMLの重要機能(1)

- NewsMLはマルチメディア対応
  - テキスト、画像、音声、動画…**将来追加されるフォーマットまで対応**
    - システム間で対応するフォーマットを決める
  - NewsMLは個々のフォーマットを包むニュース・フォーマット
- 豊富な付加情報と拡張性でニュースのライフサイクルを全サポート
  - サブジェクトコード、優先度、取り消し、改版、対象者…
  - 独自に付加情報を追加可能



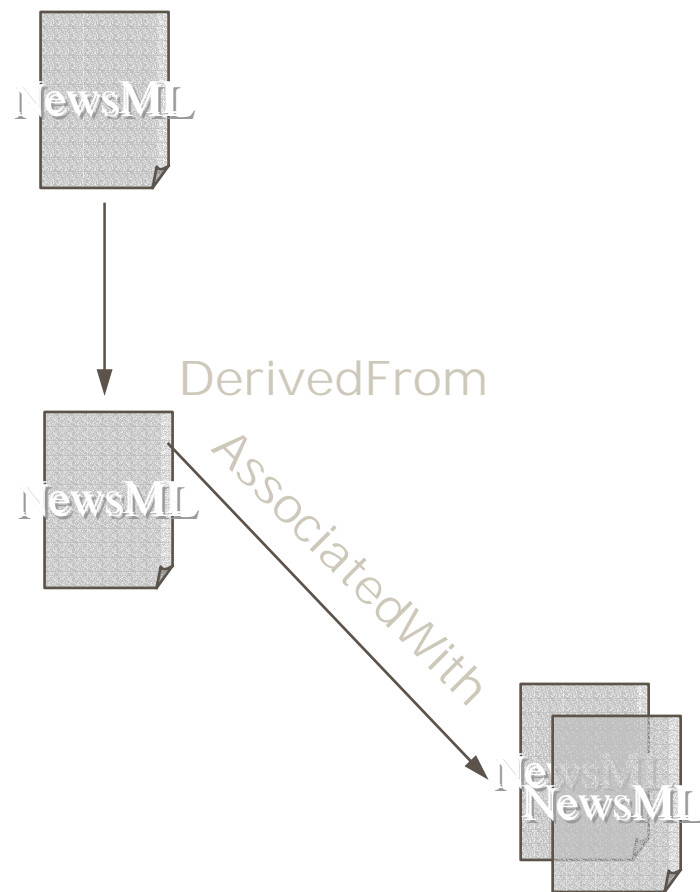
# NewsMLの重要機能(2)

- ユニークIDとリビジョン管理機能
  - すべてのNewsML/NewsItemは世界で唯一(ユニーク)なIDであるNewsML-URNを持つ
    - `urn:newsml:ibm.com:20010927:jp-sv001-hzabcdefg...:4`
  - リビジョン番号もNewsML-URNの一部
    - 省略時は最新のリビジョンを示す



# NewsMLの重要機能(3)

- リンク情報
  - NewsML-URNで関連ファイルの指定が可能
    - 派生元(DerivedFrom)
    - 関連先(AssociatedWith)



# NewsMLアプリケーション 開発時の考慮点

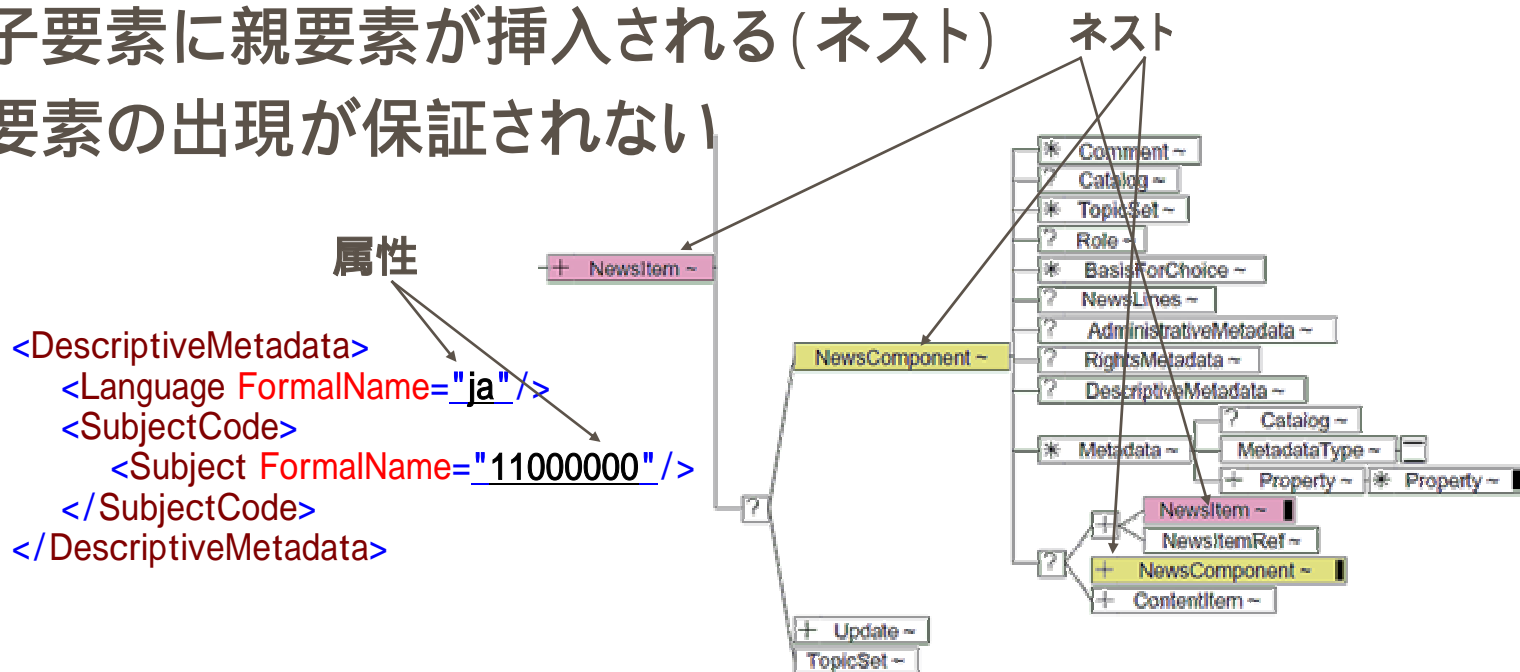




# NewsMLアプリケーション 開発時の考慮点(1)

## ■ NewsMLはXMLの中でも難しい？

- 属性を多用している
- 子要素に親要素が挿入される(ネスト)
- 要素の出現が保証されない



# NewsMLアプリケーション 開発時の考慮点(2)



- NewsMLはXMLの中でも難しい？
  - 共通辞書であるTopicSetは外部ファイルで定義される

```
<Resource>  
  <Urn>urn:newsml:iptc.org:20001006:topicset.iptc-status-ja:1</Urn>  
  <Url>../topicsets/topicset.iptc-status-ja.xml</Url>  
  <DefaultVocabularyFor Context="Status/@FormalName" />  
</Resource>
```

- DefaultVocabularyForの仕組み
- NewsMLが正しく動くエディタはXMLSpy
  - <http://www.xmlspy.com/>

# NewsMLアプリケーション 開発時の考慮点(3)

- NewsMLの妥当性チェック
  - IPTC提供のNewsMLチェッカー
    - DTD妥当性
    - 外部ファイルの妥当性
    - FormalNameの妥当性
    - NewsMLのすべての仕様を満たしているわけではない。
      - 日本IBMで改良版を作成。



# NewsMLの情報サイト

- 日本新聞協会 <http://www.pressnet.or.jp/>
  - ページ上の「NSK NewsML」をクリックしてください。
    - NewsML仕様書 (IPTCが作成したNewsML仕様書の和訳)
    - NskNewsMLレベル1解説書 (日本でどのように使っていくかを書いた解説書)
    - サンプル・ファイル
- IPTC (英文) <http://www.iptc.org/>
  - 左の小さい「NEWSML」をクリックしてください。下記の情報などが入手できます。  
(<http://www.iptc.org/site/NewsML/NewsMLSpec.htm>)
    - 英文の仕様書
    - NewsMLのDTD
    - DTDのツリー図
    - NewsMLチェッカー
- NewsMLを紹介するWEBサイト (英文) <http://www.newsml.org/>
- 毎日新聞 <http://www.mainichi.co.jp/>
  - NewsML紹介 <http://www.mainichi.co.jp/digital/newsml/>
  - WEBページではNewsMLのソースを公開しています。トップ記事の末尾の青い<More>をクリックします。表示される本文の末尾の青いロゴ「MainichiNewsML」をクリックしてください。
- ロイター <http://about.reuters.com/japan/product/NewsML.html>

ご静聴ありがとうございました。

