

Textnachrichtenformat – NewsML-G2



Textnachrichtenformat – NewsML-G2

Was NewsML-G2 ist und kann:

NewsML-G2 ist ein umfassendes Datenformat für den Austausch von Nachrichten und anderen Informationen im Nachrichtengewerbe und ist ein Produkt der IPTC (International Press Telecommunications Council – www.iptc.org), dem weltweiten Verband von Nachrichtenagenturen, anderen Nachrichtenherausgebern und Systemherstellern für die Medienbranche. Die dpa gehört der IPTC als Mitglied an.

Das Ziel von NewsML-G2 ist es, alles übermitteln zu können, was von professionellen Herausgebern von Nachrichten verbreitet wird:

- Journalistische Nachrichten in den gängigen Medientypen Text, Bild, Grafik, Audio und Video
- Selektierte und strukturierte Sammlungen von solchen journalistischen Nachrichten, in NewsML-G2 als Pakete bezeichnet.
- Strukturierte Daten über Ereignisse/Termine.
- Daten über die redaktionelle Planung der aussendenden Nachrichtenredaktion („Welche Art von Nachricht in welchem Umfang bis wann auf welchem Dienst geliefert wird.“)
- Die Definition und weltweit eindeutige Kennung von Begriffen, die als Stichworte, Klassifikationen oder sachliche Referenzen des Nachrichteninhalts verwendet werden.
- Ganze Sammlungen (auch Vokabulare genannt) solcher Begriffe

Textnachrichtenformat – NewsML-G2

- Unterstützung aktueller Standards bei der Verarbeitung
 - XML, HTML5, Semantic Web
- Unterstützung des Lifecycle von Medienobjekten
 - planned, embargoed, withheld, usable, update, correction, canceled
- Verknüpfung von Planung und Produktion
 - Termin und Planung wird zum gruppierenden Anker mehrerer Objekte
- Medien- und Darstellungsneutral
 - Einheitliches (Meta-) Datenmodell über alle Medientypen
 - z. Zt.: Text, Bild, Termin, Planung, Pakete (geplant: Audio, Video, Grafik)
 - Transformation des Contents in beliebige Darstellungsformen
- Verknüpfung beliebiger Medienobjekte
 - z. B. Text / Bild – Verknüpfung, Erzeugung von Paketen

Textnachrichtenformat – NewsML-G2 - Prozessierbare Metadaten

- **Dienst(e)**

```
<service qcode="dpasrv:ndk">
  <name role="nrol:full">Nachrichtendienst Kinder</name>
  <name role="nrol:mnemonic">ndk</name>
</service>
```

- **(Entstehungs-)Ort der Meldung**

```
<located>
  <name>Eltville</name>
</located>
```

- **Qualifiziertes und sprachunabhängiges Genre**

```
<genre qcode="dpatextgenre:14">
  <name xml:lang="de">KORR-Bericht</name>
</genre>
```

- **Qualifiziertes und sprachunabhängiges Ressort**

```
<subject qcode="dpacat:sp" type="dpatype:category">
  <name role="nrol:mnemonic" xml:lang="de">sp</name>
</subject>
```

- **Links verweisen auf**
 - externe Quellen
 - von dpa publizierte Objekte

```
<link href="http://dpaq.de/JE7tF" rank="1" rel="irel:seeAlso">
  <title>Infos zum Hörbuch</title>
</link>
<link residref="urn:newsml:dpa.com:20090101:130813-99-06655" rank="1" rel="irel:associatedWith">
  <type qcode="ninat:picture"/>
  <title>Das Hörbuch</title>
</link>
```

- **Subjects sind**

- sprachunabhängig
- immer voll qualifiziert aus einer verwalteten Taxonomie
- tragen Beziehungen zu anderen Taxonomien
- tragen untereinander einen Rang

```
<subject  
<subject qcode="dpasubject:306" rank="2" type="dpatype:dpasubject">  
  <name xml:lang="de">Immobilien</name>  
  <sameAs qcode="medtop:20000241">  
    <name xml:lang="de">Immobilien</name>  
  </sameAs>  
</subject>
```

- **Keywords sind**

- natürlich sprachliche Begriffe
- können als Hashtag qualifiziert werden
- tragen untereinander einen Rang

```
<keyword rank="1">Steuern</keyword>
```

```
<keyword rank="2" role="drol:hashtag">#BTW</keyword>
```

Textnachrichtenformat – NewsML-G2 - Qualifizierung von Geografika

```
<assert literal="ID0EKE">
  <POIDetails>
    <position latitude="50.9600263" longitude="6.9266830"/>
    <address>
      <line role="nrol:display">Rheinfighters</line>
      <line role="nrol:full">Arnimstraße 2, 50825 Köln, Deutschland</line>
      <locality>
        <name>Köln</name>
      </locality>
      <country qcode="dpacountry:1" type="cpnat:geoArea">
        <name role="nrol:display" xml:lang="de">Deutschland</name>
        <sameAs qcode="iso3166-1a2:DE"/>
        <sameAs qcode="iso3166-1a3:DEU"/>
        <broader qcode="wldreg:r150">
          <name xml:lang="en-GB">Europe</name>
        </broader>
      </country>
    </address>
  </POIDetails>
</assert>
```


Textnachrichtenformat – NewsML-G2 - XHTML Auszeichnung des Contents

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Merkels Japan-Reise und eine Kernfrage </title>
    <meta charset="utf-8"/>
  </head>
  <body>
    <header>
      <ul class="slugline">
        <li class="subject" data-qcode="medtop:20000638">Internationale Beziehung</li>
        <li class="geo" data-qcode="dpacountry:1">Deutschland</li>
        <li class="geo" data-qcode="dpacountry:70">Japan</li>
      </ul>
      <time class="publicationDate" data-datetime="2015-03-08T06:00:04+01:00">08.03.2015 06:00</time>
      <p class="genre dpatextgenre-14">KORR-Bericht</p>
      <h1>Merkels Japan-Reise und eine Kernfrage </h1>
      <p class="byline">Von Kristina Dunz und Lars Nicolaysen</p>
    </header>
    <section class="main dpatextgenre-14">
      <p class="teaser">Viel Zeit hat die Kanzlerin nicht in Japan. Ein Tag Tokio und dann noch ein A
        <span class="dateline">Berlin/Tokio <span class="credit">(dpa)</span> - </span>Die Physiker:
      </p>
    </section>
  </body>
</html>
```

Textnachrichtenformat – NewsML-G2 - XHTML Auszeichnung des Notizblocks

```
<edNote role="dpaednoterole:notepad">
  <section xmlns="http://www.w3.org/1999/xhtml" class="notepad">
    <header>
      <h3>Notizblock</h3>
    </header>
    <section class="np-public">
      <h4>Internet</h4>
      <ul>
        <li>
          <a href="http://dpaq.de/ac100">Webseite des japanischen Regierungschefs</a>
        </li>
      </ul>
      <h4>Orte</h4>
      <ul>
        <li>[Merkel-Besuch] (Präfektur Tokio, Japan)</li>
        <li>[Bundeskanzleramt] (Willy-Brandt-Straße 1, 10557 Berlin, Deutschland)</li>
      </ul>
    </section>
    <section class="np-nonpublic">
      <p class="np-constraint">Die folgenden Informationen sind nicht zur Veröffentlichung bestimmt</p>
      <h4>Ansprechpartner</h4>
      <ul>
        <li>Bundespresseamt: +49 30 18 272 2030</li>
      </ul>
    </section>
  </section>
</edNote>
```