

{Scheme, Code} tuples for News Industry Taxonomies

22 May 2006

Quick intro to NewsML 2

- NewsML is *the* standard for B2B News Exchange
- Developed by the International Press Telecommunications Council (IPTC) -- <http://www.iptc.org>
- 62 members: AFP, ANA, ANSA, AP, APA, BBC, Business Wire, Canadian Press, CCI, Dow Jones, dpa, EBU, EFE, HINA, IFRA, ITAR-TASS, NSK, KUNA, Kyodo News, MTI, NAA, New York Times, PA, PR Newswire, Reuters, TT, UPI, Xinhua, ...
- NewsML 2 EP#1 started Dec 2005, EP#2 started May 2006
- Completion of NewsML 2 first edition due Dec 2006

NewsML 2 examples

- `<pubStatus code="stat:usable"/>`
- `<locCreated code="city:Paris"/>`
- `<creator code="team:DOM"/>`
- `<subject code="cat:04000000"/>`
- `<subject code="isin:NL0000361939"/>`
- `<subject code="pers:021147"/>`
- `<description role="dsc:caption">Plusieurs centaines de salariés du BTP (2.000 selon les organisateurs, 700 selon la police) manifestent, le 15 mai 2005 à Paris, pour exiger ...</description>`

Requirements

- Scheme and Code maintained as distinct objects
- Both Scheme and {Scheme, Code} map to dereferenceable URIs
- A very compact representation of {Scheme, Code} tuples
- Codes be the real-World values, including numerics
- Strong preference for colon notation (ie Scheme:Code)
- Scheme alias ↔ Scheme URI mappings
 - provider dependent
 - can be remote and cached

IPTC NewsML 2 Catalog

<catalog>

<ns prefix="acdc" uri="http://iptc.org/newscodes/audiocodec#" />

<ns prefix="cosp" uri="http://iptc.org/newscodes/colourspace#" />

<ns prefix="gre" uri="http://iptc.org/newscodes/genre#" />

<ns prefix="oito" uri="http://iptc.org/newscodes/ofinterestto#" />

<ns prefix="prv" uri="http://iptc.org/newscodes/provider#" />

<ns prefix="rol" uri="http://iptc.org/newscodes/role#" />

<ns prefix="scn" uri="http://iptc.org/newscodes/scene#" />

<ns prefix="pst" uri="http://iptc.org/newscodes/pubstatus#" />

<ns prefix="suc" uri="http://iptc.org/newscodes/subjectcode#" />

IPTC NewsML 2 Catalog (cont)

```
<ns prefix="vcdc" uri="http://iptc.org/newscodes/videocodec#" />
<ns prefix="cnt" uri="http://iptc.org/newscodes/contenttype#" />
<ns prefix="drl" uri="http://iptc.org/newscodes/descriptionrole#" />
<ns prefix="rdn" uri="http://iptc.org/newscodes/rendition#" />
<ns prefix="sgn" uri="http://iptc.org/newscodes/edsignal#" />
<ns prefix="trl" uri="http://iptc.org/newscodes/titlerole#" />
<ns prefix="cpt" uri="http://iptc.org/newscodes/concepttype#" />
<ns prefix="why" uri="http://iptc.org/newscodes/whypresent#" />
<ns prefix="irl" uri="http://iptc.org/newscodes/itemrelation#" />
</catalog>
```

Status of Scheme:Code notation

- Deployed by Reuters for three years with NewsML 1
- Included in NewsML 2 specifications for nearly a year
- Being tested by IPTC members and others as part of NewsML 2 EP#1 and EP#2
- GRDDL transforms to RDF/XML and N3 under development
- Working with W3C RDF-in-XHTML TF to achieve synergy in our handling of tuples
- Working with W3C Multimedia Semantic Incubator group to create an OWL model of NewsML 2

More info available from

- <http://www.iptc.org>
- <http://groups.yahoo.com/group/newsml-2>
- misha.wolf@reuters.com