RDF Validation requirements for data about products, services and companies

Mark Harrison (University of Cambridge)
Robert Beideman (GS1 Global Office)

mark.harrison@cantab.net
robert.beideman@gs1.org
Global not-for-profit community organization developing user-driven open standards since 1973:
- Bar codes and product identifiers, RFID
- Supply chain visibility
- Product master data exchange, product categories

Global network of research labs working together since 1999/2000 and with GS1 in development of:
- Technology to support an “Internet of Things”
- Supply chain visibility, security, efficiency
- Next generation networks, sensors, etc.
- MIT, University of Cambridge, ETH Zurich / St Gallen, Keio University, KAIST, Fudan University, University of Adelaide
GS1 Digital:

Using technology such as Linked Open Data (RDF etc.) to make trustworthy data about products and services available on the web
To improve visibility of products and suppliers
To enable enhanced consumer interactivity with products and support new services / applications
LOD for products – Pre-Sale

1. User types in a **keyword to specify type of product**

2. Keyword is matched in GPC hierarchy and appropriate **contextual filters** are displayed to refine the search

3. **Additional constraints** can be specified by the user e.g. available **budget / price range**, **urgency** (how long to wait?) and **proximity** (how far away?)

---

**SEARCH**

- **Milk**
  - Organic?
  - Skimmed
  - Half Fat
  - Full Fat

- **Max. Price**: £1
- **How Urgent?**: 1 hr
- **How Far?**: 1 km

**Search for Products & Services**

- **Tesco Organic Semi-Skimmed Milk**
  - £0.58
  - < 1 km away
    - (10 mins walk)
Global Product Classification as LOD
LOD for products – Post-Sale

URL (user manual)
URL (user forum)
URL (product support)
URL (firmware updates)
URL (driver software)
URL (warranty registration)
Sourcing of structured master data
Sourcing of structured master data

GS1 Global Product Classification (GPC)
+ attribute - value pairs

Product Master Data
Party Master Data
in GDSN / GS1 Source
EU Food Labelling legislation

- **EU 1169/2011 Regulation** includes Article 14 on **Distance Selling**
- Mandatory details to be freely available to consumers before purchase
- Grocery websites will typically need to provide info about
  - Ingredients
  - Potential allergens
  - Nutritional information
- Legislation does not require this information to be machine-readable
- However, EU 1169/2011 is driving a major compliance effort to make this kind of information available on manufacturers’ / retailers’ websites – and there is an **opportunity** to make it available as linked data at the same time, to open up much richer search applications for food products.
Liability issues

- Data needs to be correct, up-to-date, trustworthy
- Units of measure need to be precise, unambiguous
  e.g. grams per 100g / 100ml of product
  (also micrograms, milligrams per 100g / 100ml)

- Changes to product formulation (ingredients) could introduce new allergens (risk of death) or make a product no longer suitable for those on specific diets (vegan, kosher, etc.)
## Technical validation requirements

<table>
<thead>
<tr>
<th>Type of requirement</th>
<th>Example for product master data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardinality constraints / Completeness</td>
<td>Are all mandatory attributes specified? (with non-null values) Which attributes are optional?</td>
</tr>
<tr>
<td>Cardinality constraints within complex data structures such as Quantitative Values</td>
<td>Are required units of measure specified precisely?</td>
</tr>
<tr>
<td>Range checking for codified values. Values of specific attributes are members of a</td>
<td>Units of measure or product categories, attributes and values take code values or URIs from a defined code list or controlled vocabulary, where expected.</td>
</tr>
<tr>
<td>defined code list or controlled vocabulary</td>
<td></td>
</tr>
<tr>
<td>Checking for current reachability of online resources (when the Object of an RDF</td>
<td>Use RDF validation to check for broken hyperlinks (e.g. RDF links to product manuals etc.)</td>
</tr>
<tr>
<td>triple is a URL)</td>
<td></td>
</tr>
<tr>
<td>Canonicalization of a specified ordered set of RDF triples in order to use digital</td>
<td>Manufacturers might publish master data about products and retailers might replicate this data in their websites. Potential liability issues from inaccurate product information if a product’s ingredients list changes and introduces an allergen / unacceptable ingredient</td>
</tr>
<tr>
<td>signatures to detect any deviations from a corresponding set of RDF triples from</td>
<td></td>
</tr>
<tr>
<td>a source that is considered to be authoritative</td>
<td></td>
</tr>
</tbody>
</table>
Maybe not for validating source data

- Authoritative Master Data about products, services, companies
  - Export as XML
    - Transform to an XML subset relevant for consumers (e.g. via XSLT or Xquery)
    - Publish on web as XML data
  - Validate (via XSD)

- Transform to RDF or HTML + RDFa (via XSLT or Xquery templates)
  - Web pages containing RDFa markup of master data

- Use W3C Provenance standards to point to data source and transformations

Snippet Generators / Data export tools available for use by Brand Owners, Retailers
e.g. in GDSN or GS1 Source

© 2013 GS1
Improve markup / validation tools

Demo at http://www.autoidlabs.org.uk/RDFa-test/markup-rdfa.html

RDF Visualization: http://rdfa.info/play
Promote consistent convergent structured data through ‘profiling’

Which vocabularies? Which terms?

Promoting correct markup for complex values

Quantitative Value
- weight
- value: 600
- unitCode: “GRM”
Support dynamic contextual search interfaces

Cardinality / Enumeration constraints can drive UI options for contextual search

Boolean choice

Choose 1 from N
Questions / Discussion