PROV-O profile
for describing and tracking editorial actions on Controlled Vocabularies

NKOS 2014 London

Jean Delahousse, Consultant, Infeurope
Johan de Smedt, Chief Technology Officer, TenForce
Agis Papantoniou, Senior Project Manager, TenForce
Background

The project derives from existing and ongoing work for the team in charge of the EU controlled vocabularies [4] [5] and for the team in charge of the content and metadata repository "CELLAR" at the Publications Office of the EU.

The project is done with the support and advices from:

– Marc Kuster
– Willem van Gemert
– Christine Laaboudi
– Madeleine Kiss
– Luc Bertrand
Summary

- Context
- Changes management today
- New needs, new services
- Project and Tasks
- Q&A, How we can make it a more collective project?
- References
Controlled vocabularies are evolutionary intellectual work (see presentation about DEWEY evolutions by A. Slavic & A. Isaac [3])

Evolutions should be traced to understand the rationales of the changes, remember the timing and identify the sources

Evolutions should be formalised and precisely described
  - For the users of the KOS to be aware of the changes and calculate the impact on their information system
  - For the machines which use the KOS to automatically apply the changes and keep trace of them
Changes management today

1. Context

- EU Publications Office edits and disseminates several controlled vocabularies used by information systems of the EU institutions and by external organisations.

- The team uses a formal process to apply changes to control vocabularies but also to disseminate the new versions and the description of the changes.

- The EU Publications Office is generalizing the dissemination of the controlled vocabulary using LOD standards (DCAT and SKOS).
Changes management today

2. A JIRA ticket logging a business demand

[Image of JIRA ticket]

**CORP - Create new code for National Banks "BANK_NAT"**

- **Type:** Implementation
- **Priority:** Minor
- **Affects Version/s:** None
- **Component/s:** AT corporate-body
- **Labels:** None
- **Add to Release no...** Yes
- **Impact assessment:** Minor
- **Revision Excel:** 10388
- **Revision XML:** 10380
- **Release notes:**

  Creation of the following concepts for lists of national banks within and outside the Eurosystem:
  - BANK_EURO11
  - BANK_EURO12
  - BANK_EURO13
  - BANK_EURO15
  - BANK_EURO16
  - BANK_EURO17
  - BANK_EURO18
  - BANK_NONEURO4
  - BANK_NONEURO3
  - BANK_NONEURO13
  - BANK_NONEURO14
  - BANK_NONEURO12
  - BANK_NONEURO11
  - BANK_NONEURO10
Changes management today

3. Adding one of the concepts in the editorial application
Changes management today

4. XML description of the changes for a single business demand (fragment)
Changes management today

5. Excel publication of the xml file recap of the changes for one business demand

<table>
<thead>
<tr>
<th>XML Diff report for</th>
<th>corporate-body</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report #id:</td>
<td></td>
</tr>
<tr>
<td>Generated:</td>
<td>20140520 16:12:26</td>
</tr>
<tr>
<td>Previous [current.entries.date]</td>
<td>2014-05-20</td>
</tr>
<tr>
<td>Current [current.entries.date]</td>
<td>2014-05-20</td>
</tr>
<tr>
<td>Numbers of &quot;Concepts&quot; modified</td>
<td>13</td>
</tr>
<tr>
<td>Numbers of &quot;Records&quot; modified</td>
<td>13</td>
</tr>
<tr>
<td>&quot;Update&quot; operations:</td>
<td>38</td>
</tr>
<tr>
<td>&quot;Insert&quot; operations:</td>
<td>0</td>
</tr>
<tr>
<td>&quot;Delete&quot; operations:</td>
<td>12</td>
</tr>
<tr>
<td>Creation of new CAT code</td>
<td>0</td>
</tr>
<tr>
<td>Deprecation of a CAT code (code should no longer be used, but remains in list)</td>
<td>0</td>
</tr>
<tr>
<td>Deletion of a CAT code (Must be avoided, retired instead)</td>
<td>0</td>
</tr>
<tr>
<td>Addition of a CAT label (prefLabel)</td>
<td>0</td>
</tr>
<tr>
<td>Addition of a CAT label (altLabel)</td>
<td>0</td>
</tr>
<tr>
<td>Modification of an existing CAT label (prefLabel)</td>
<td>0</td>
</tr>
<tr>
<td>Modification of an existing CAT label (altLabel)</td>
<td>0</td>
</tr>
<tr>
<td>Addition of a code mapping</td>
<td>0</td>
</tr>
<tr>
<td>Deletion of a code mapping</td>
<td>0</td>
</tr>
<tr>
<td>Admin Status changed</td>
<td>12</td>
</tr>
</tbody>
</table>
Changes management today

6. Excel publication of the xml file
detail on the insertion of one Concept

<table>
<thead>
<tr>
<th>ID</th>
<th>authority-code</th>
<th>adm.status</th>
<th>Insert /Upd.</th>
<th>old</th>
<th>new</th>
<th>element</th>
<th>xpath</th>
</tr>
</thead>
<tbody>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>BANK_EURO11</td>
<td>authority-code</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/authority-code</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>EU</td>
<td>corporate.classification</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/corporate.classification</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>2000-12-31</td>
<td>end.use</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/end.use</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>0</td>
<td>label</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/label</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>EUMS</td>
<td>parent.id</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/parent.id</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>1999-01-01</td>
<td>start.use</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/start.use</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>IMMC</td>
<td>use.context</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/use.context</td>
<td></td>
</tr>
<tr>
<td>COB0008113</td>
<td>BANK_EURO11</td>
<td></td>
<td>Insert element</td>
<td>National banks within the Euros...</td>
<td>label[1]/lg[@lg='eng']</td>
<td>/corporatebodies/record[id=&quot;COB0008113&quot;]/label[1]/lg[@lg='eng']</td>
<td></td>
</tr>
</tbody>
</table>
Changes management today
7. The JIRA ticket for the business demand documented with the detailed description on the changes (xml, html, xls)
Changes management today
8. Recap of the workflow

Change demand
- E-mail from an EU institution

JIRA ticket
- Describe the change to be made and related sources

Apply changes
- Editorial tool

Detailed description
- XML describing the impact of one business change
- xls, html publication
- Update of the JIRA ticket

Compilation of changes between Rel N to Rel N+1
- xml
- Excel
- PDF
- HTML
New needs – new services

Business needs

• The dissemination of the changes between versions should be improved to answer new business needs:

  – Disseminate the information about the rationales and sources of the changes (more transparency)

  – Disseminate the information about the changes using LOD standards instead of proprietary format (facilitate access to information)

  – Allow the CELLAR (content management system of the EU Publications Office based on RDF) to describe the changes of the controlled vocabulary and make them available to end users (facilitate access to information)

  – Create advanced services to facilitate access to legal content when the KOS used for their semantic annotation has changed over time (facilitate access to information, more transparency)
New needs – new services
What needs to change?

• The business demand, recorded in the JIRA ticket is not formalized and published for dissemination

• The details of the changes applied to the vocabulary are not related to the original demand
  – it can be tricky to understand the details of the change without description of their rationales
  – The audit trace is not disseminated

• The formal description of the changes are not disseminated using Linked Open Data format, whereas the vocabulary is

• The formal description of the changes cannot be directly processed by a RDF repository to update a controlled vocabulary based on RDF/SKOS
New needs – new services

- E-mail from an EU institution

- Describe the change to be make and related sources
- Formalize the business demand with RDF based on the PROV ontology based on JIRA demand description

- Editorial tool

- XML describing the impact of one business change
- xls, html publication
- rdf publication based on PROV / SKOS
- Update of the JIRA ticket

- rdf/PROV/SKOS description of business demand, sources and detailed vocabulary changes
- Excel, PDF, HTML
Project & Tasks

• Objective:
  – Create a PROV application profile to manage changes on controlled vocabulary

• Restriction:
  – The PROV application profile will be done for a SKOS serialisation of the controlled vocabulary
Project & Tasks

• Task 1: formalisation of the Jira ticket

  – PROV profile to describe the business demand, the rationales, the sources

  – Taxonomy of controlled vocabulary changes from a business point of view
    • Create a list of concepts
    • Add a translation to all the concepts for this language
    • Move a branch
    • Deprecate a concept
Project & Tasks

• Task 2: formalize the detailed changes applied to the KOS
  – PROV profile to describe changes at concept / property level
  – Taxonomy of controlled vocabulary changes at a concept level and property level
    • Create a new concept E as narrower concept of F
    • Deprecate concept D
    • Change the Croatian pref label of concept C
    • Merge a concept A in the concept B
  – Open discussion
    • Relation between a description of the changes related to the business view of the vocabulary (as handle by the back office) and a description of the changes related to the SKOS representation of the vocabulary
Project & Tasks

• Task 3: Creation of test datasets
  – Release N of the MDR Corporates Vocabulary (xml/rdf dataset based on euvoc application profile)
  – Release N+1 of the MDR Corporates Vocabulary
  – Dataset describing the version changes between release N and N+1 based on the PROV application profile
Project & Tasks

• Task 4: Test in the CELLAR environment
Project & Tasks

- Task 4: Test in the CELLAR environment
Project & Tasks

• Task 4: Test in the CELLAR environment
Task 4: Test in the CELLAR environment
Q&A

How to make it a more collective project?

Contact: delahousse.jean@gmail.com
References

acronyms

- KOS: Knowledge Organization System
- MDR: Metadata Registry
- LOD: Linked Open Data
Possible PROV-O mapping

- Activity represents the JIRA issue
- Entity the concept and related properties and resources affected
Possible PROV-O mapping

- Time boundaries of the entity
- Change details
- Entity details