

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 on going 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

Context

- 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

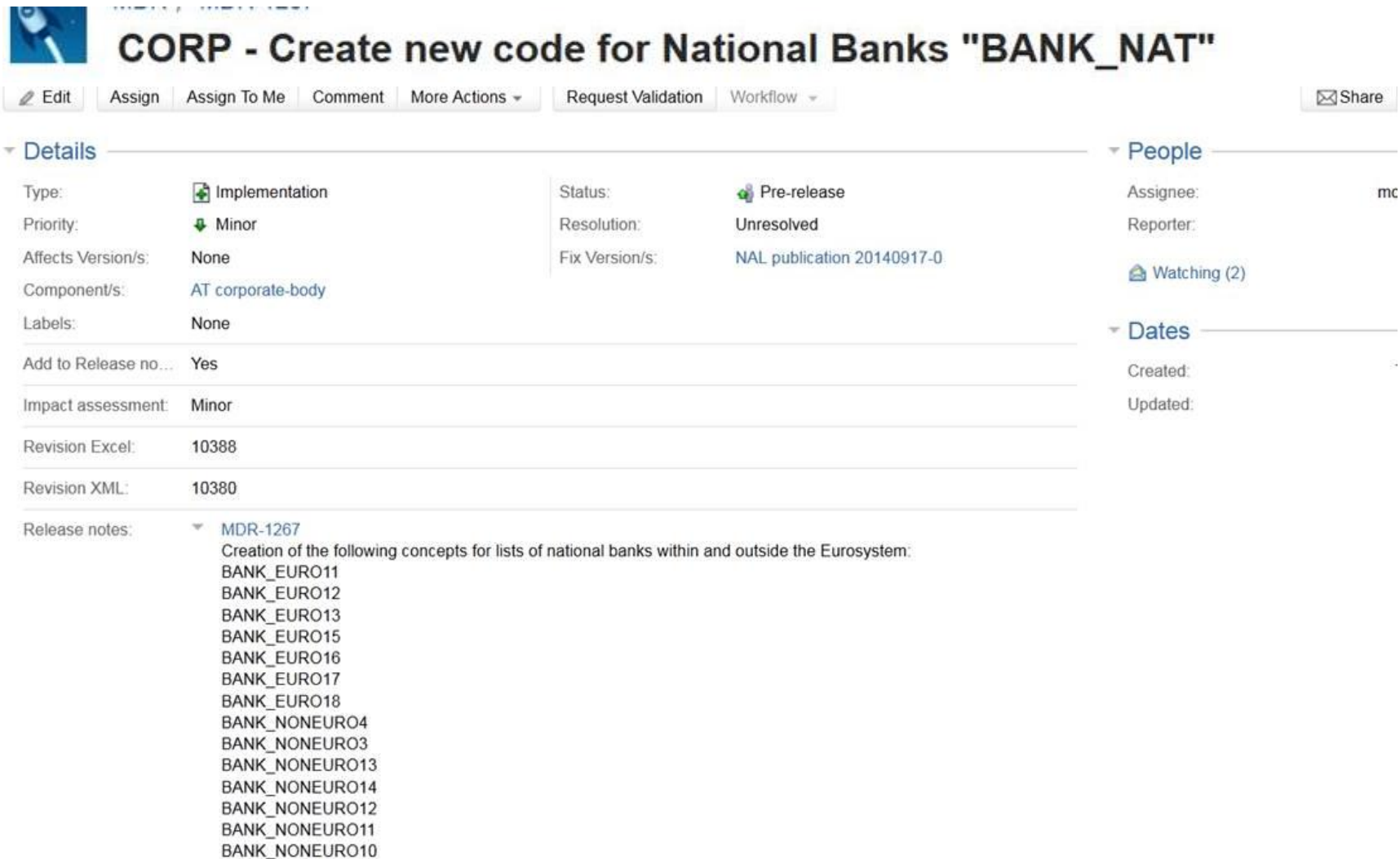
Changes management today

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



The screenshot shows a JIRA ticket interface. At the top, there is a blue header with a logo on the left and the ticket title "CORP - Create new code for National Banks \"BANK_NAT\"". Below the title is a navigation bar with buttons for "Edit", "Assign", "Assign To Me", "Comment", "More Actions", "Request Validation", and "Workflow". A "Share" button is located on the right side of the navigation bar.

The main content area is divided into two columns. The left column is titled "Details" and contains the following information:

- 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: MDR-1267
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

The right column is titled "People" and contains the following information:

- Assignee: mc
- Reporter:
- Watching (2)
- Dates: Created: Updated:

Changes management today

3. Adding one of the concepts in the editorial application

The image consists of three screenshots of a data management application interface, showing the process of adding a concept. Red arrows and circles highlight key elements.

Top Screenshot: Shows a table with columns: authority-co, admin_stat, eng_label, parent_id, acronym, protocol_or, classificati, creation_d, start_use_d, end_use_d, predecesso, successor, LOC_id. The row for ID 0812 shows 'BANK_EURO1' under authority-co, 'deprecated' under admin_stat, 'National banks' under eng_label, and 'EU' under classificati. The status bar at the bottom shows 'COB' circled in red, with 'op-mapped-code' highlighted in green. Red arrows point from the 'COB' circle to the 'authority-co', 'eng_label', and 'classificati' columns.

Middle Screenshot: Shows a table with columns: ID, authority-co, bul_label, spa_label, ell_label, eng_label, fra_label, gle_label. The row for ID COB0000812 shows 'BANK_EURO11' under authority-co and 'National banks within the Eurosystem:' under eng_label. The status bar at the bottom shows 'COB' circled in red, with 'labels' highlighted in yellow. Red arrows point from the 'labels' circle to the 'eng_label' and 'fra_label' columns.

Bottom Screenshot: Shows a table with columns: ID, authority-co, admin_status, eng_label_ignore, COMREF, DEMED, EurLex, IMMC, OJ, PUB_CORP. The row for ID COB0000812 shows 'BANK_EURO11' under authority-co, 'deprecated' under admin_status, and 'National banks within the Eurosystem:' under eng_label_ignore. The status bar at the bottom shows 'COB' circled in red, with 'use_context' highlighted in orange. Red arrows point from the 'use_context' circle to the 'COMREF', 'DEMED', 'EurLex', 'IMMC', and 'OJ' columns.

Changes management today

4. xml description of the changes for a single business demand (fragment)

```
▼<concept authority-code="EP_AGRI">
  ▼<record record_id="COB0000802">
    ▼<context xpath="/corporatebodies/record[1]/successor[2]" src="from-file" record_id="COB0000802"
      authority-code="EP_AGRI" adm.status="deprecated" deprecated="true" IMMC.approval.date="2014-09-05"
      use_context.immc="true">
      ▼<element name="successor" isComplexType="false" occurrence="1" context-string="successor">
        ▼<character_differences>
          ▼<comment>
            The element_value method handles those cases where the text data differs between two
            elements that have the same name, namespace URI, and are at the same location in the
            document tree. Note that all whitespace is normalized and the text from mixed-content
            elements (those containing both text and child elements mixed together) is aggregated down
            to a single value.
          </comment>
          <old_value>COB0000801</old_value>
          <new_value>COB0000271</new_value>
        </character_differences>
      </element>
    </context>
  </record>
</concept>
```


Changes management today

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

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

Changes management today

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

	A	B	C	D	E	F	G	H	
1	ID	authority-code	adm.status	Insert /Upda	old	new	element	xpath	xpath4impac
57	COB0000813	BANK_EURO11	candidate.depre	insert element		BANK_EURO11	authority-code	/corporatebodies/record[@id='COB0000813']/authority-	/corporatebodi
58	COB0000813	BANK_EURO11	candidate.depre	insert element		EU	corporate.classificati	/corporatebodies/record[@id='COB0000813']/corporate	/corporatebodi
59	COB0000813	BANK_EURO11	candidate.depre	insert element		2000-12-31	end.use	/corporatebodies/record[@id='COB0000813']/end.use	/corporatebodi
60	COB0000813	BANK_EURO11	candidate.depre	insert element		o	label	/corporatebodies/record[@id='COB0000813']/label	/corporatebodi
61	COB0000813	BANK_EURO11	candidate.depre	insert element		EUMS	parent.id	/corporatebodies/record[@id='COB0000813']/parent.id	/corporatebodi
62	COB0000813	BANK_EURO11	candidate.depre	insert element		1999-01-01	start.use	/corporatebodies/record[@id='COB0000813']/start.use	/corporatebodi
63	COB0000813	BANK_EURO11	candidate.depre	insert element		IMMC	use.context	/corporatebodies/record[@id='COB0000813']/use.conte	/corporatebodi
64	COB0000813	BANK_EURO11	candidate.depre	insert element		National banks within the Eurosystem: Nationale Bank	lg.version[@lg='eng']	/corporatebodies/record[@id='COB0000813']/label[1]/lg	/corporatebodi

D	E	F	G
Insert /Upda	old	new	element
insert element		BANK_EURO11	authority-code
insert element		EU	corporate.classificati
insert element		2000-12-31	end.use
insert element		o	label
insert element		EUMS	parent.id
insert element		1999-01-01	start.use
insert element		IMMC	use.context
insert element		National banks within the Eurosystem: Nationale Bank	lg.version[@lg='eng']

Changes management today

7. The JIRA ticket for the business demand documented with the detailed description on the changes (xml, html, xls)



CORP - Create new code for National Banks "BANK_NAT"

- Edit
- Assign
- Assign To Me
- Comment
- More Actions ▾
- Resolve Issue
- Close Issue
- Workflow ▾

Click to add description

Attachments

corporate-body-xmldiff-2014-05-20-with-2014-05-20.html	297 kB	20/05/14 16:29
corporate-body-xmldiff-2014-05-20-with-2014-05-20.html	127 kB	20/05/14 16:25
corporate-body-xmldiff-2014-05-20-with-2014-05-20.xlsx	16 kB	20/05/14 16:29
corporate-body-xmldiff-2014-05-20-with-2014-05-20.xlsx	11 kB	20/05/14 16:25

Issue Links

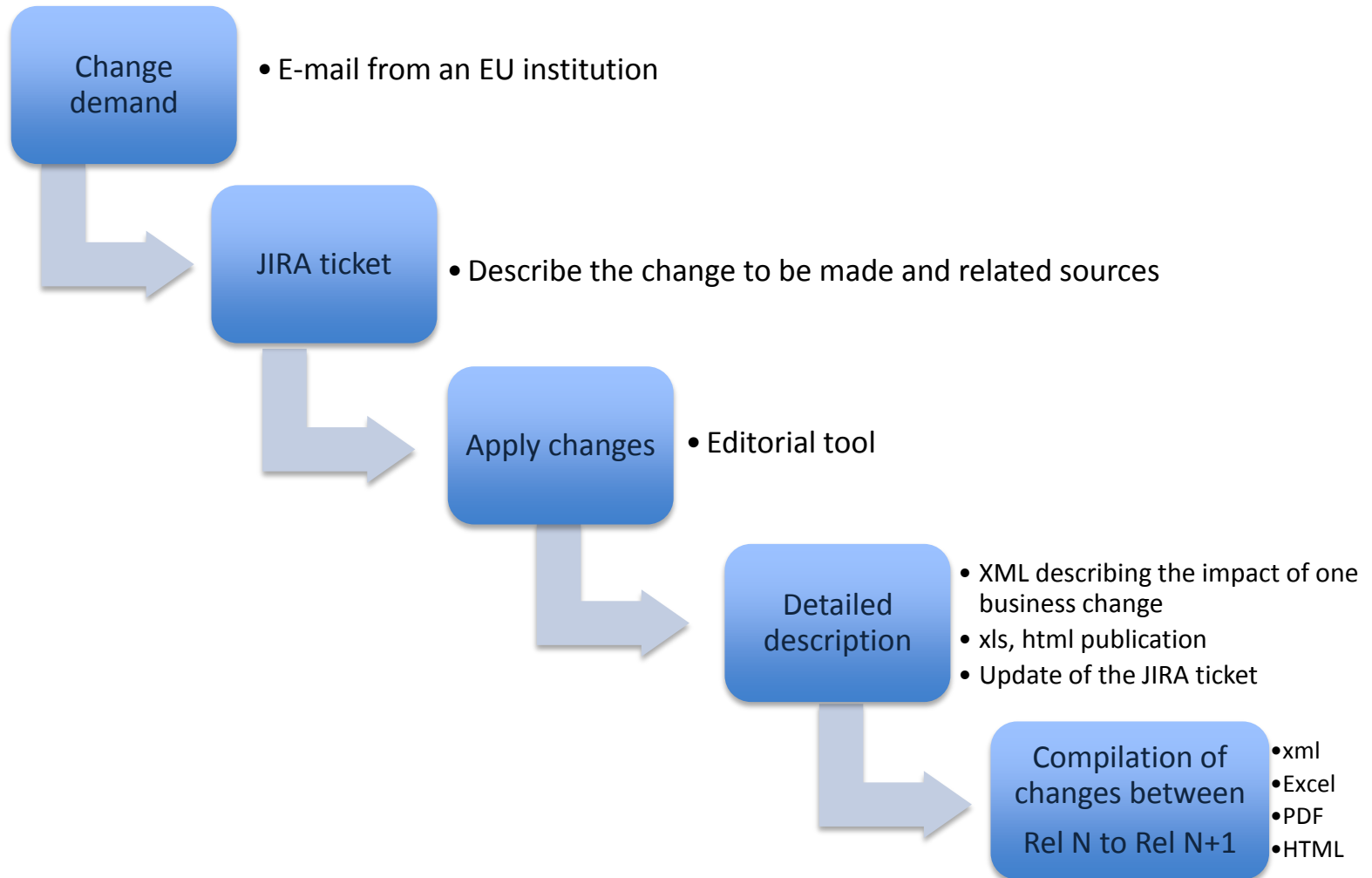
includes	MDR-1265 CORP - add the mapping code EIGE from the corresponding FD_110 table	
is included in	MDR-1406 CORP - Corrections from Style Guide to be implemented	
relates to	MDR-1639 Corporate body AT: Start impact assessment for publication nal-20140917-0	

Activity

All Comments Work Log History Activity Transitions Subversion Commits

Changes management today

8. Recap of the workflow



New needs – new services

Business needs

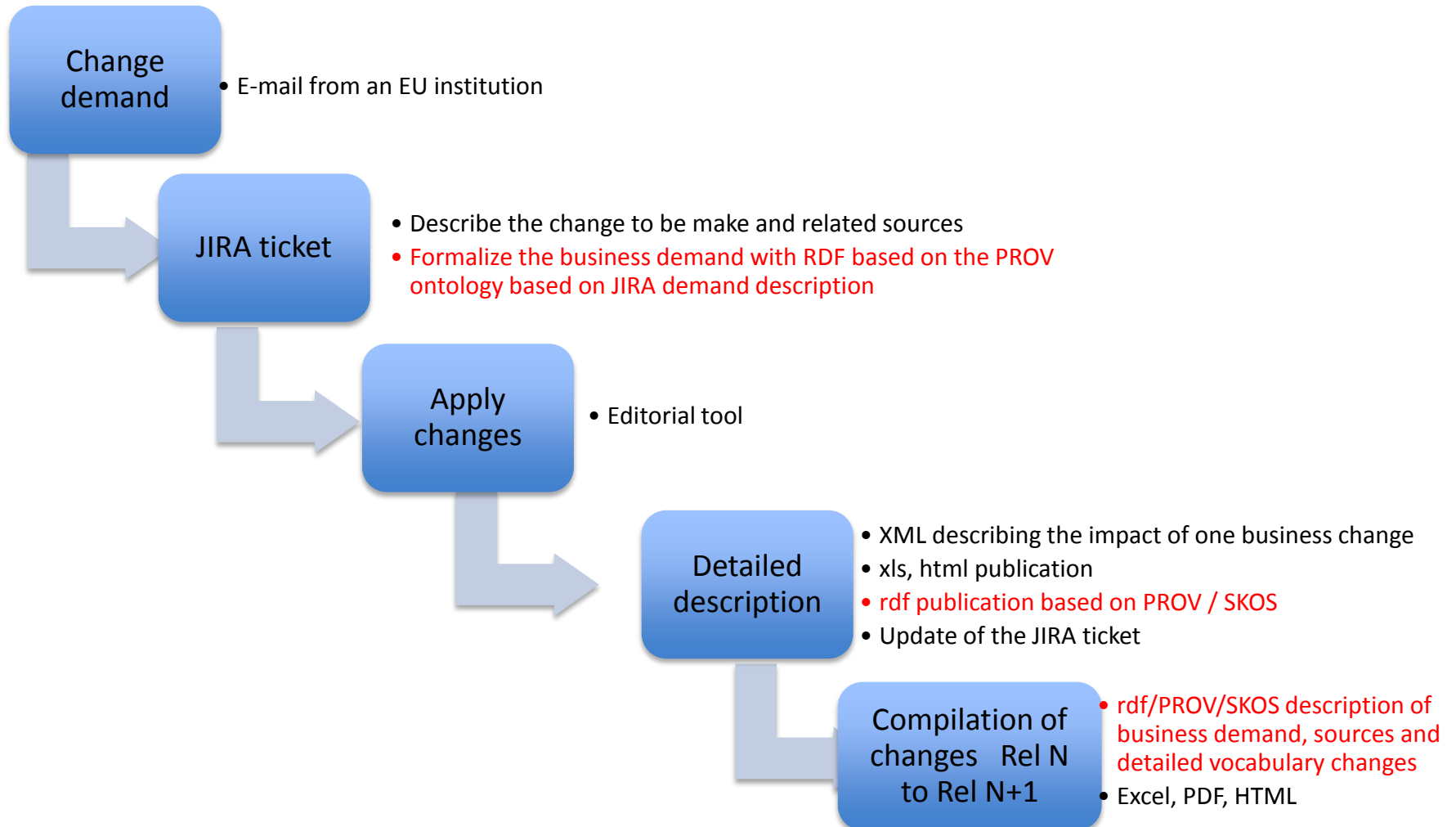
- The dissemination of the changes between versions should be improved to answer new business needs:
 - Disseminate the information about the rationales and sources of the changes (more transparency)
 - Disseminate the information about the changes using LOD standards instead of proprietary format (facilitate 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



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

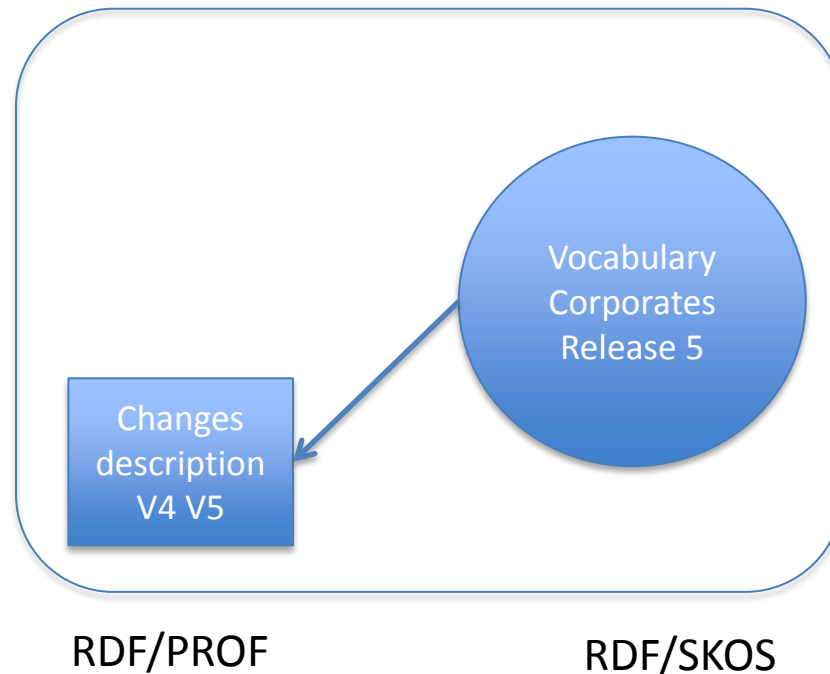
- Task 2: formalize the detailed changes applied to the KOS
 - PROV profile to describe changes at concept / property level
 - Taxonomy of controlled vocabulary changes at a concept level and property level
 - Create a new concept E as narrower concept of F
 - Deprecate concept D
 - Change the Croatian 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

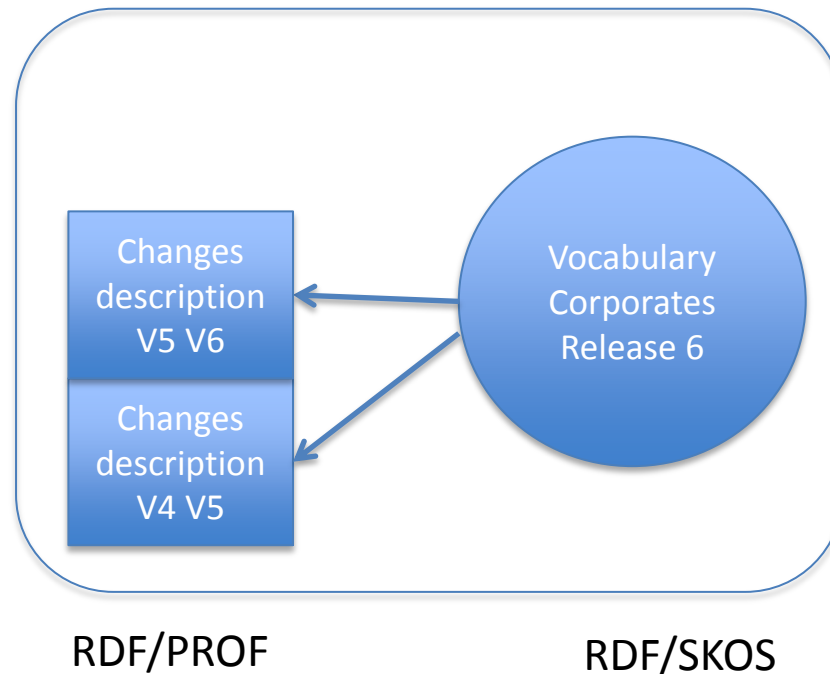
Project & Tasks

- Task 4: Test in the CELLAR environment



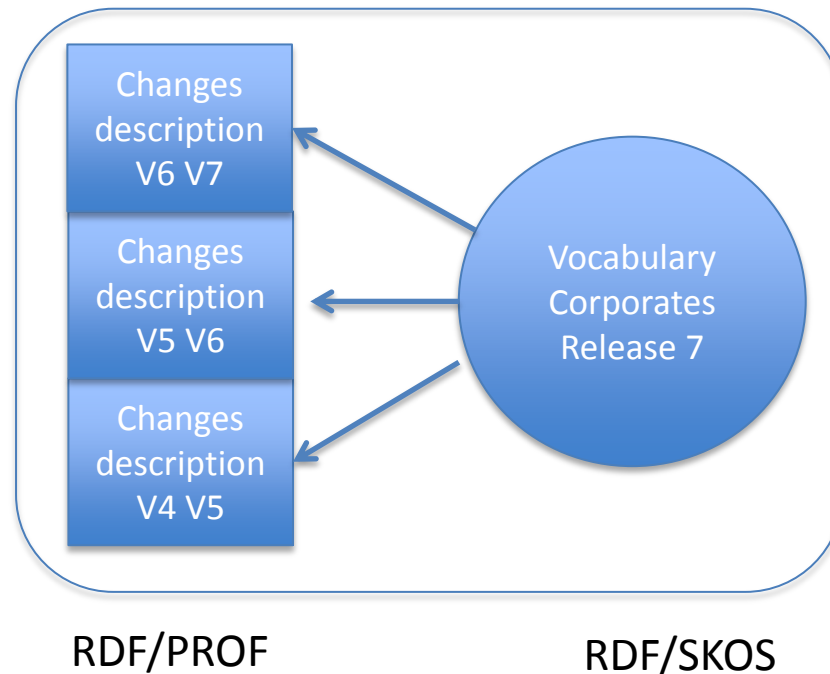
Project & Tasks

- Task 4: Test in the CELLAR environment



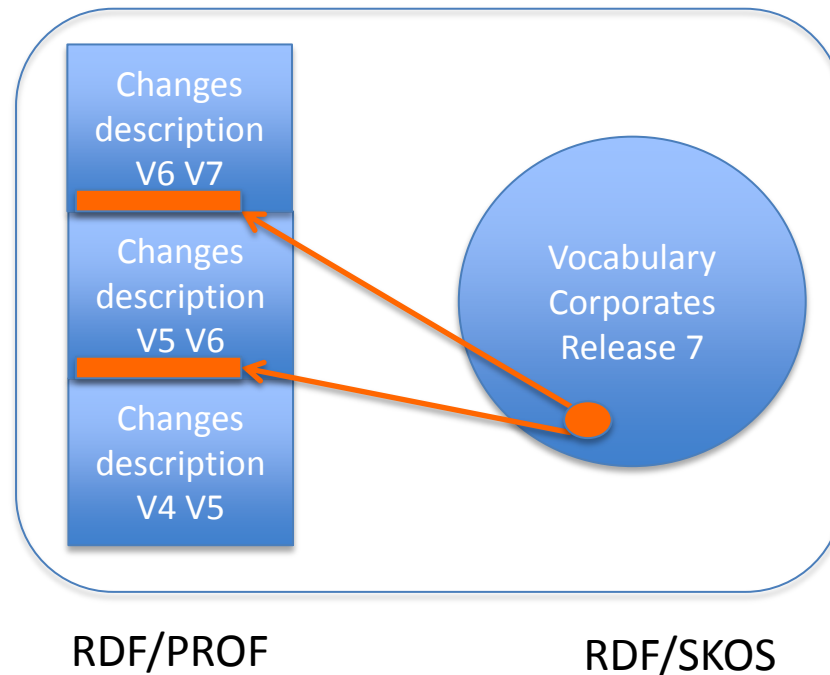
Project & Tasks

- Task 4: Test in the CELLAR environment



Project & Tasks

- Task 4: Test in the CELLAR environment



Q&A

How to make it a more collective
project ?

Contact : delahousse.jean@gmail.com

References

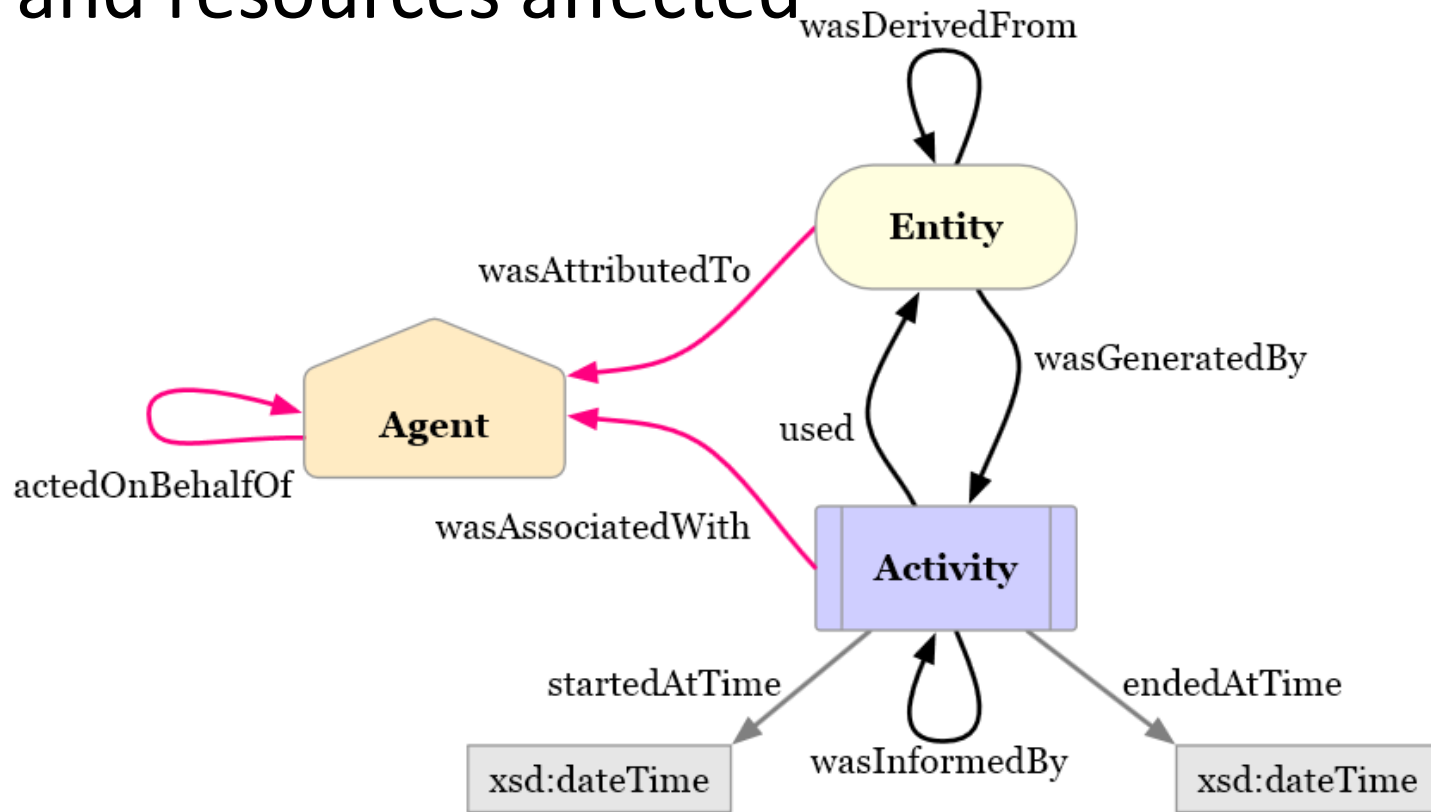
- [1] <http://www.w3.org/2001/sw/wiki/SKOS/Issues/ConceptEvolution>, Retrieved 04/07/2014
- [2] PROV-O ontology, <http://www.w3.org/TR/prov-o/>, Retrieved 04/07/2014
- [3] Slides on change management in Identifying management issues in networked KOS: examples from classification schemes. Aida Slavic, Antoine Isaac.
<http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2009/programme.html>
- [4] Publications Office of the EU, Metadata Registry portal
<http://publications.europa.eu/mdr/authority/>
- [5] Publications Office of the EU, Eurovoc portal, <http://eurovoc.europa.eu/>

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

