

Constantly Under Construction - STW Thesaurus for Economics Linked Data Maintenance

Joachim Neubert

ZBW – Leibniz Information Centre for Economics

DCMI Libraries Community

DC-2013, Lisbon, Portugal

05.09.2013

STW Thesaurus for Economics

- Created in the 1990s, currently maintained and enhanced by ZBW
- More than 6,000 descriptors in English and German
- Fine-meshed network of concepts:
 - More than 16,000 broader/narrower relations
 - 11,000 related concepts
 - Additional access path through a term classification with more than 500 entries

STW versions and changes against the previous

8.04 – 29.2.2009

- 223 new descriptors (skos:Concept with skos:prefLabel de/en)
- 4 deleted descriptors
- 618 new non-descriptors (skos:altLabel)

8.06 – 22.4.2010

- 58 new descriptors
- 57 deleted descriptors
- 491 new non-descriptors

8.08 – 30.6.2011

- 105 new descriptors
- 141 deleted descriptors
- 1490 new non-descriptors

8.10 – 21.3.2012

(Counts stem from the legacy maintenance system and may not be as accurate as they seem, but should give an idea about the change frequency)

Other kinds of changes

Other kinds of changes may be highly relevant especially for the creation and evolvement of mappings from or to other vocabularies:

- Changes of the preferred label
„Partnership“ -> „Business partnership“
- Changes of alternative labels
„Bankkonkurs“ -> „Bankenkonkurs“
- Changes of relations (broader/narrower/related)
- Labels attached to new concepts
„Industrial cluster “ from „Geographic concentration “ (deleted) to
„Regional Cluster “
- Splits and merges of concepts

How do we currently handle this?

RDF statements about a particular version

```
<http://zbw.eu/stw>  
  a skos:ConceptScheme, void:Dataset ;  
  dcterms:issued "2012-03-21"^^xsd:date ;  
  owl:versionInfo "8.10" ;  
  ...
```

STW URI versioning concept

Stable URIs for skos:ConceptScheme und skos:Concept

- <http://zbw.eu/stw>
- <http://zbw.eu/stw/descriptor/19664-4>

303 redirect to versioned URLs (RDFa files, when required with language extension)

- <http://zbw.eu/stw/versions/latest/about>
- <http://zbw.eu/stw/versions/latest/descriptor/19664-4/about>

Archived RDFa/rdf/ttl files available (with hint to the latest version)

- <http://zbw.eu/stw/versions/8.06/about>
- <http://zbw.eu/stw/versions/8.06/descriptor/19664-4/about>

Search functions and web services always work on the latest version

Pragmatic solution – overview page

STW Thesaurus for Economics

Versions

Prior versions of the STW are provided here for reference ([Changes](#)).

Published versions have even version numbers. Odd version numbers are reserved for internal purposes.

- [8.10](#) (Detailed Changelog - in German)
- [8.08](#) (Detailed Changelog - in German)
- [8.06](#) (Detailed Changelog - in German)
- [8.04](#) (first web and linked data version)

Changes are traceable only intellectually (but at all)

Please use the language- and version-independent URIs to link to the concepts (eg. <http://zbw.eu/stw/descriptor/19664-4> instead of <http://zbw.eu/stw/versions/latest/descriptor/19664-4/about.en.html>).

Detailed changelog

From legacy maintenance system (simple text file, in German):

Neuangelegte Deskriptoren:

1. Aborigines (Australien) [engl.: Aboriginal Australians] (26584-4)
2. Afghanen [engl.: Afghans] (26068-1)
3. Afghanisch [engl.: Afghan] (26069-6)
4. Afrikaans [engl.: Afrikaans] (26070-0)
5. Afrikaner [engl.: Africans] (26071-5)
6. Afrikanisch [engl.: African] (26072-3)
7. Albaner [engl.: Albanians] (26082-0)
8. Albanisch [engl.: Albanian] (26083-5)
9. Amerikaner [engl.: Americans] (26084-3)
10. Amerikanisch [engl.: American] (26085-1)
11. APEC-Staaten-seitig [engl.: From APEC countries] (26086-6)
12. Araber [engl.: Arabs] (26101-1)
13. Arabisch [engl.: Arab] (26102-6)
14. Armenier [engl.: Armenians] (26103-4)
15. Aserbaidshanisch [engl.: Azerbaijani] (26129-0)
16. Asiaten [engl.: Asians] (26130-1)
17. Asiatisch [engl.: Asian] (26131-6)
18. Ausländisch [engl.: Foreign] (26132-4)
19. Austauschtheorie (Soziologie) [engl.: Social exchange theory] (25974-3)

Deleted descriptors/concepts

URI is still defined – shown on a RDFa page like this:

Real estate loan

*Deprecated (used at last in version 8.04), USE **Mortgage***

```
<http://zbw.eu/stw/descriptor/12257-3>
  a skos:Concept, zbwext:Descriptor ;
  skos:inScheme <http://zbw.eu/stw> ;
  rdfs:label "Real estate loan"@en, "Realkredit"@de ;
  owl:deprecated true ;
  dcterms:isReplacedBy <http://zbw.eu/stw/descriptor/13775-4> ;
  skos:historyNote "Deprecated (used at last in version
8.04) "@en .
```

How to handle this better?

What users want to know when we publish a new version:

- What's new?
- What has changed?

Rough plan for a SKOS history

- 1) Create a raw diff of sorted n-triple files. (This gives you thousands and thousands of differences, even excluding bnodes.)
- 2) Group changes for each concept.
- 3) Recognize insertion and deletion of concepts as a whole (presumably the most important changes).
- 4) Recognize certain types of changes (e.g., altered prefLabel, added altLabel, changed relations).
- 5) Enrich the concept URIs with the preferred label (in a given language).

Rough plan for a SKOS history (continued)

- 6) Arrange everything nicely on a RDFa overview page
(additions/deletion of concepts, perhaps some of the more important types of changes, statistics such as amount of changed/unchanged concepts, etc.)
- 7) Provide a change history per concept, e.g. on a RDFa page which can be linked from a concept page.
- 8) Optionally, if the terminology includes meta-structures such as a subject classification, add aggregated information about the most intensively changed subject areas (hot spots) to the overview page.

URIs for versions and version deltas

URI for version, e.g.

- `http://zbw.eu/stw/version/8.04`

URI for delta, e.g.

- `http://zbw.eu/stw/version/8.04/delta/8.06`

Each delta has a deletions and a insertions part, e.g.

- `http://zbw.eu/stw/version/8.04/delta/8.06/insertions`
- `http://zbw.eu/stw/version/8.04/delta/8.06/deletions`

named graphs



History of a SKOS concept

Implementation prototype based on version and delta files, loaded as named graphs into a SPARQL endpoint. Example CONSTRUCT result:

```
<http://zbw.eu/stw/descriptor/10112-4>
  skos:prefLabel "Welfare analysis"@en , "Wohlfahrtsanalyse"@de ;
  :history [ :altLabelDeletion "Wohlfahrtsmessung"@de , "Wohlfahrtsmaß"@de , "Welfare measurement"@en , "Wohlstandsmaß"@de ;
    :altLabelInsertion "Welfare effect"@en , "Wohlfahrtsgewinn"@de , "Wohlfahrtseffekt"@de , "Wohlfahrtsverlust"@de ;
    :delta <http://zbw.eu/stw/version/8.08/delta/8.10> ;
    :prefLabelDeletion "Welfare effect"@en , "Wohlfahrtseffekt"@de ;
    :prefLabelInsertion "Wohlfahrtsanalyse"@de , "Welfare analysis"@en
  ] .
```

(Very preliminary implementation, currently limited to prefLabel and altLabel changes)

Thanks

Joachim Neubert

ZBW – Leibniz Information Centre for Economics

j.neubert@zbw.eu

<http://zbw.eu/stw>

<http://zbw.eu/labs>