

Leveraging SKOS to Trace the Overhaul of the STW Thesaurus for Economics

Joachim Neubert

ZBW – Leibniz Information Centre for Economics, Hamburg

Dublin Core Conference

São Paulo, Brazil

03.09.2015

STW Thesaurus for Economics

- Created in the 1990s, now maintained and enhanced by ZBW
- More than 6,000 descriptors in English and German
- Since 2009 published as Linked Open Data in SKOS
- Roughly every year a new version
- Major overhaul since 2010 – subject area by subject area
- Finished in summer 2015 with STW 9.0
 - 777 new descriptors
 - 1,052 descriptors deprecated (mostly merged into others)
 - thousands and thousands of other changes

Further outline

- General versioning approach with STW
- User questions and requirements
- Getting a grip on changes:
the dataset versioning and skos-history approach
 - Overview and application
 - Selected useful reports
- Outlook: Future work and the skos-history project

How did we handle KOS evolution from start?

RDF statements about a particular version

```
<http://zbw.eu/stw>  
  a skos:ConceptScheme, void:Dataset ;  
  dcterms:issued "2015-06-15"^^xsd:date ;  
  owl:versionInfo "9.0" ;
```

...

Others do this in a similar, yet slightly different way (dcterms:modified, dcterms:hasVersion, ...) – and sometimes, this changes over time

STW versions in URIs

Stable URIs for skos:Concept (and similar for skos:ConceptScheme)

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

redirect (303) to URLs (html+rdfa/rdf/ttl files) about the current version

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

Archived html+rdfa/rdf/ttl files available

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

Deprecated concepts

No deletion – URI is still defined, shown on a HTML/RDFa page like this:

Real estate loan

*Deprecated (used at last in version 8.04) ([link](#)),
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 .
```

What users want to know ...

... when we publish a new KOS version:

- What's new?
- What has changed?

Use cases for extended change information

- Human indexers wanting to learn about new and deprecated concepts
- Human indexers (and supporting applications) re-indexing large sets of documents
- People maintaining mappings to other vocabularies, and applications supporting them
- People maintaining a derived subset of a KOS
- Vocabulary-based automatic or semi-automatic indexing applications
- Search applications

Getting a grip on changes

Provided that we have no access to the KOS maintenance system where the changes take place originally, or can't extend it to report this changes comprehensively.

The dataset versioning + skos-history approach

=> should basically work on every SKOS vocabulary

Three basic steps to an actionable skos-history

Start with one SKOS file per version.

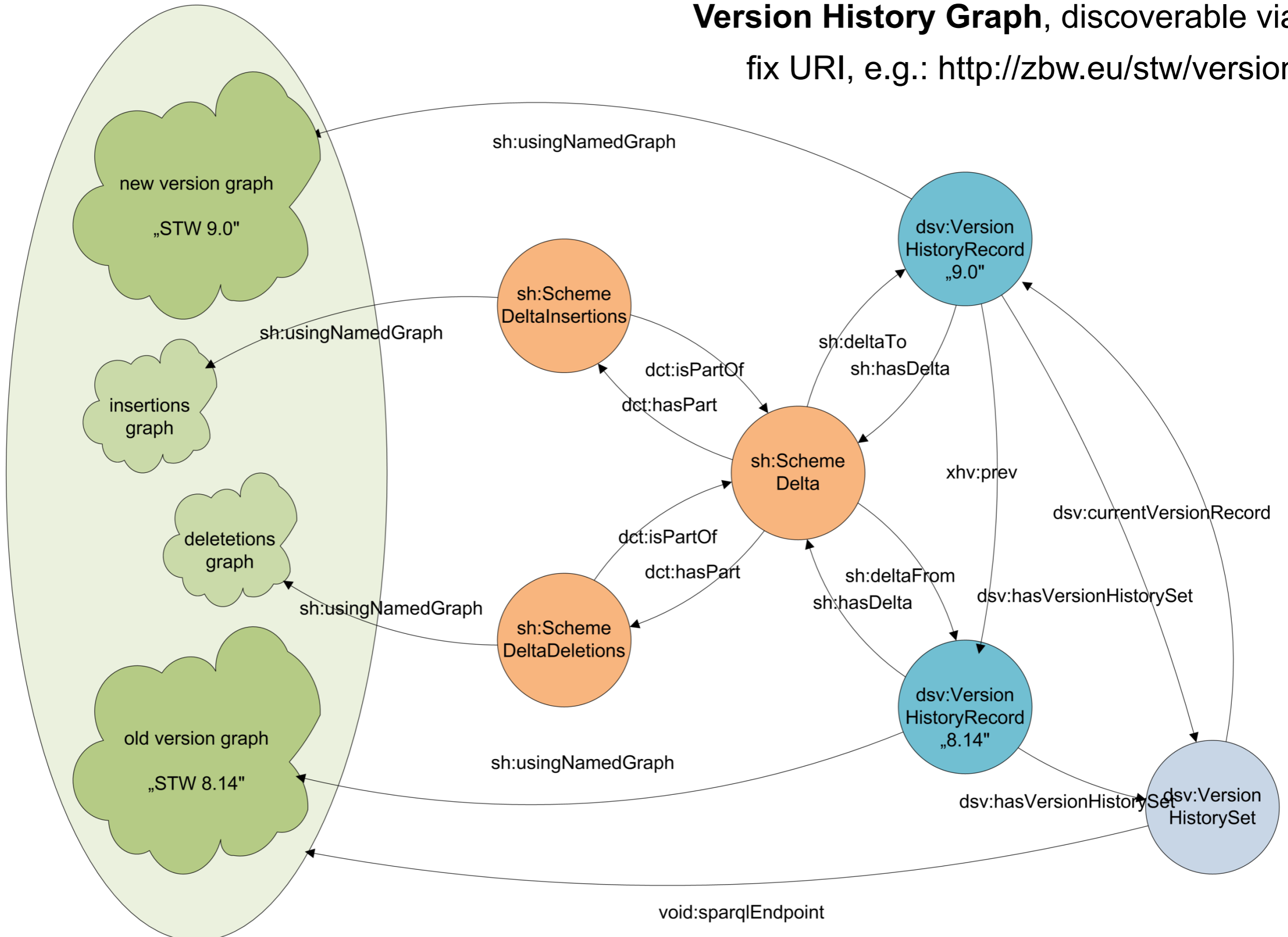
1) Create the deltas - insertions and deletions - between every two version files.

(Via a raw diff of sorted ntriples files, or via SPARQL MINUS in a triple store. This gives you thousands and thousands of differences - added or deleted triples -, even excluding bnodes.)

2) Load the version files and the insertions and deletions into a triple store as named graphs.

3) Add metadata about the versions and the deltas in a separate „version history graph“.

Version History Graph, discoverable via fix URI, e.g.: <http://zbw.eu/stw/version>



Example endpoint: <http://zbw.eu/beta/sparql/stwv/query>

Vocabularies used for the plumbing

- dc:/dcterms:
Dublin Core, as usual the base for everything
- void: <http://rdfs.org/ns/void#>
Vocabulary of interlinked datasets
- sd: <http://www.w3.org/ns/sparql-service-description#>
SPARQL service description
- delta: <http://www.w3.org/2004/delta#>
Differences between RDF graphs
- dsv: <http://purl.org/iso25964/DataSet/Versioning#>
Version history records (providing version identifier and date) and a pointer to the current version – outside the actual version data
- sh: <http://purl.org/skos-history/>
Scheme and concept version deltas

What's the benefit?

A database of all versions of a KOS and all deltas between versions
– which can be queried in parallel!

```

SELECT distinct (?concept AS ?addedConcept) (str(?prefLabel) AS ?addedConceptLabel)
WHERE {
  # parameters
  VALUES ( ?versionHistoryGraph ?language ) {
    ( <http://zbw.eu/stw/version> "en" )
  }
  GRAPH ?versionHistoryGraph {
    # the compared versions default to the current and the previous one
    ?versionset dsv:currentVersionRecord/xhv:prev/dc:identifier ?oldVersion .
    ?versionset dsv:currentVersionRecord/dc:identifier ?newVersion .
    # get the delta and via that the relevant graphs
    ?delta a sh:SchemeDelta ;
      sh:deltaFrom/dc:identifier ?oldVersion ;
      sh:deltaTo/dc:identifier ?newVersion ;
      sh:deltaFrom/sh:usingNamedGraph/sd:name ?oldVersionGraph ;
      dcterms:hasPart ?insertions .
    ?insertions a sh:SchemeDeltaInsertions ;
      sh:usingNamedGraph/sd:name ?insertionsGraph .
  }
  # for each inserted concept, a newly inserted prefLabel must exist ...
  GRAPH ?insertionsGraph {
    ?concept skosxl:prefLabel/skosxl:literalForm | skos:prefLabel ?prefLabel
  }
  # ... and the concept must not exist in the old version
  FILTER NOT EXISTS {
    GRAPH ?oldVersionGraph {
      ?concept ?p []
    }
  }
  # restrict output to a certain language
  FILTER ( lang(?prefLabel) = ?language )
}
ORDER BY ?prefLabel

```

Query for added concepts

Results: Newly inserted concepts

Results of the query:

addedConcept

- 1 [Accountants](#)
- 2 [Accounting fraud](#)
- 3 [Addiction prevention](#)
- 4 [Alternative currency](#)
- 5 [Anti-discrimination law](#)
- 6 [Asset-market approach of the exchange rate](#)

Version history page, links to ...

STW Thesaurus for Economics

Versions

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

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

The [current version](#) of STW is 9.0.

All published versions:

- [9.0 \(2015-06-15\)](#)
Change Reports: [Text](#) (in German) | [Interactive](#)
- [8.14 \(2014-11-18\)](#)
Change Reports: [Text](#) (in German) | [Interactive](#)
- [8.12 \(2013-10-30\)](#)
Change Reports: [Text](#) (in German)
- [8.10 \(2012-03-21\)](#)
Change Reports: [Text](#) (in German)
- [8.08 \(2011-06-30\)](#)
Change Reports: [Text](#) (in German)
- [8.06 \(2010-04-22\)](#)
Change Reports: [Text](#) (in German)
- [8.04 \(2009-02-16\)](#)
Change Reports: (first web and linked data version)

Change tracking via skos-history

Changes are traceable only intellectually (but at all)

... pre-compiled interactive change reports (regular since v 9.0)

STW Thesaurus for Economics

Change details in v 9.0

The interaktive change reports were generated from the Linked Data version history of STW.

Descriptors:

- Added
- Deprecated (with replacements)
- Changed preferred labels
- Added labels
- Deleted labels
- Added narrower relationships
- Added broader relationships
- Splits: Labels moved to new descriptors
- Merges: Labels moved from deprecated descriptors

Subject categories:

- Added
- Deprecated (with replacements)
- Changed notations/labels

Selected useful reports

- Changed prefLabels (descriptors) and notations (subject categories)
- Splits and merges of concepts
- Visualizing change with aggregated data
- History of a single concept

Changed notations (of STW subject categories)

old	new	concept
B.02.02.01	B.02.03	http://zbw.eu/stw/thsys/71044
B.06.02	B.06.03	http://zbw.eu/stw/thsys/70310
B.06.03	B.06.02	http://zbw.eu/stw/thsys/70471
B.11	B.01.07	http://zbw.eu/stw/thsys/179318
N.04.04.01	N.04.04.05	http://zbw.eu/stw/thsys/73365
N.04.04.02	N.04.04.06	http://zbw.eu/stw/thsys/73364
N.04.04.04	N.04.04.02	http://zbw.eu/stw/thsys/73362
N.05.08	N.05.07	http://zbw.eu/stw/thsys/73333
N.05.08.01	N.05.07.01	http://zbw.eu/stw/thsys/73332
N.05.08.02	N.05.07.02	http://zbw.eu/stw/thsys/73331

http://zbw.eu/beta/sparql-lab/?queryRef=https://api.github.com/repos/jneubert/skos-history/contents/sparql/changed_notations.rq

Merges and splits of concepts

... can be recognized by tracing the movement of labels

New concepts, split from old ones

Labels moved to added concepts:

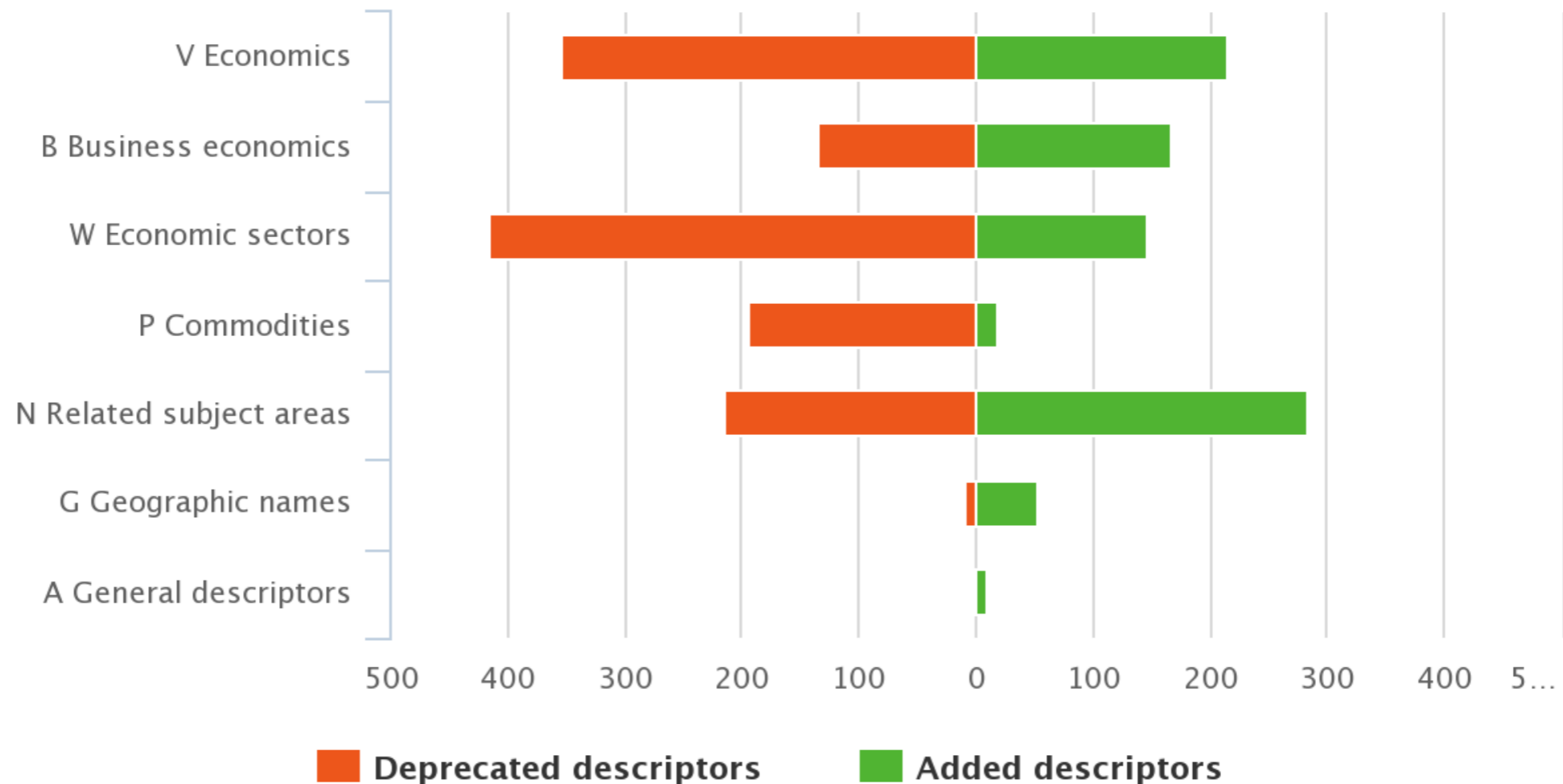
oldConcept	movedLabel	newConcept
Commercial professions	Buchhalter	Accountants
Commercial professions	Accountants	Accountants
Free-money theory (S. Gesell)	Freigeld	Alternative currency
Free-money theory (S. Gesell)	Komplementärgeld	Alternative currency
Free-money theory (S. Gesell)	Regiogeld	Alternative currency
Free-money theory (S. Gesell)	Regionalgeld	Alternative currency
Free-money theory (S. Gesell)	Regionalwährung	Alternative currency
Free-money theory (S. Gesell)	Schwundgeld	Alternative currency
Labour market discrimination	Affirmative action	Anti-discrimination law

http://zbw.eu/beta/sparql-lab/?queryRef=https://api.github.com/repos/jneubert/skos-history/contents/sparql/labels_moved_to_added_concepts.rq

Visualizing change with aggregated data

Added and deprecated descriptors (by sub-thesaurus)

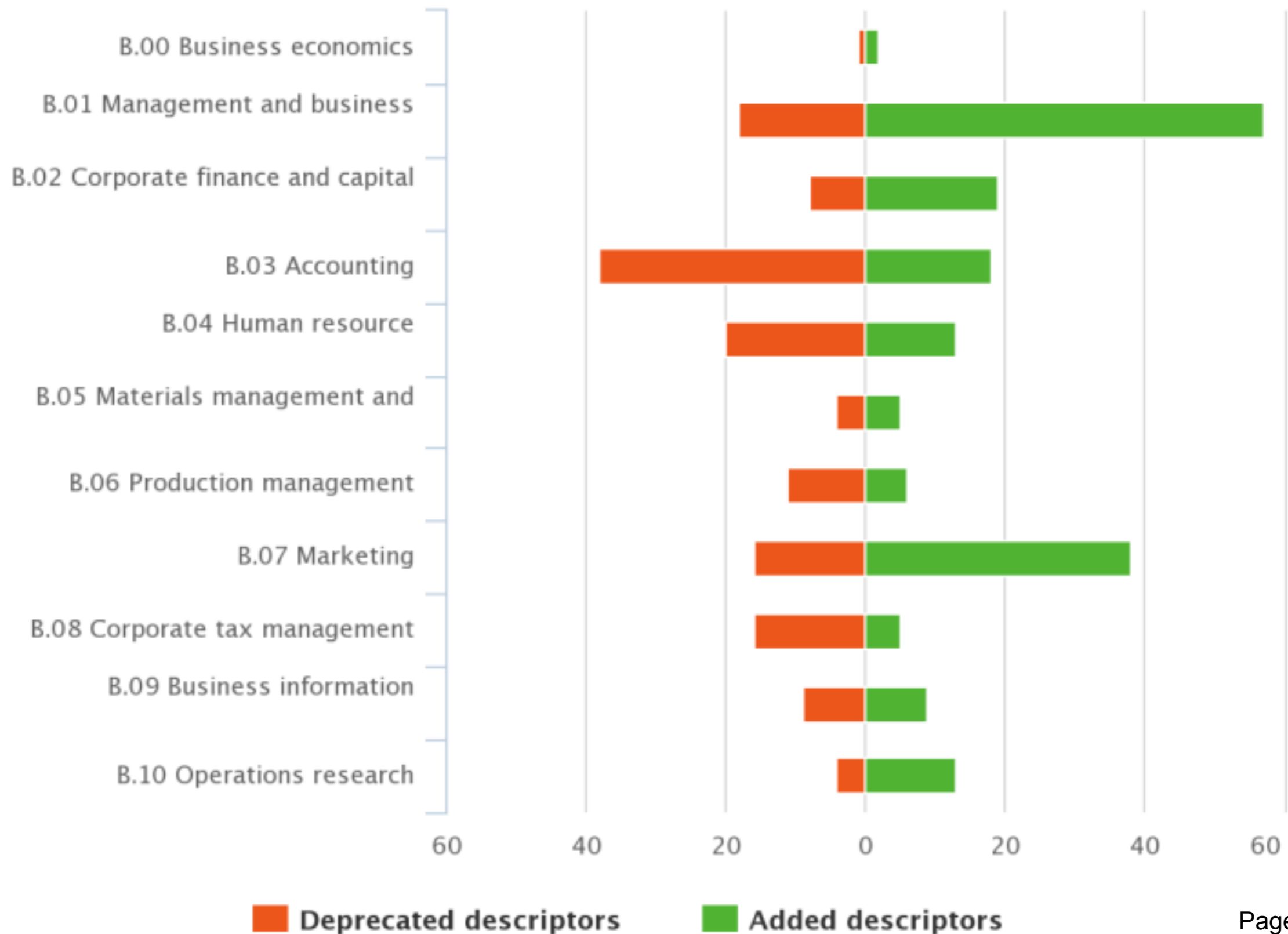
Version 8.06 to 9.0



Highcharts.com

Added and deprecated descriptors "B" (by 2nd level category)

Version 8.06 to 9.0



Drill down from chart to change report

STW Thesaurus for Economics

Added descriptors (in v 8.06 - 9.0)



Showing 1 to 38 of 38 entries (filtered from 980 total entries)

Show entries

Search:

	secondLevelCategory	addedConcept
1	B.07 Marketing	Brand architecture
2	B.07 Marketing	Brand extension
3	B.07 Marketing	Brand loyalty
4	B.07 Marketing	Brand name
5	B.07 Marketing	Cause-related marketing
6	B.07 Marketing	Celebrity endorsement
7	B.07 Marketing	Competitive analysis

Change history of a particular concept

RDF/XML RDF/Turtle **Concept history (RDF/Turtle)** Deutsch

Personnel selection EB  

Personalauswahl (german)

used for: Employment interview, Employee selection

Change history of a concept: “Personnel selection”

```
<http://zbw.eu/stw/descriptor/12571-4/version/8.06/delta/8.08>
  a sh:ConceptDelta ;
  dcterms:isPartOf <http://zbw.eu/stw/version/8.06/delta/8.08> ;
  delta:deletion [] ;
  delta:deletion [] ;
  delta:deletion [ skos:broader <http://zbw.eu/stw/thsys/70244> ] ;
  delta:deletion [] ;
  delta:deletion [ skos:altLabel "Job matching"@en ] ;
  delta:deletion [] ;
  delta:deletion [] ;
  delta:deletion [] ;
  delta:insertion [ skos:broader <http://zbw.eu/stw/descriptor/29001-4> ] ;
  delta:insertion [ skos:altLabel "Eignungsdiagnostik"@de ] ;
  delta:insertion [ skos:altLabel "Bewerberauswahl"@de ] ;
  delta:insertion [ skos:broader <http://zbw.eu/stw/thsys/180783> ] ;
  delta:insertion [ skos:related <http://zbw.eu/stw/descriptor/11189-6> ] ;
  delta:insertion [] ;
  delta:insertion [ skos:related <http://zbw.eu/stw/descriptor/15787-1> ] ;
  delta:insertion [] .

<http://zbw.eu/stw/descriptor/12571-4/version/8.08/delta/8.10>
  a sh:ConceptDelta ;
  dcterms:isPartOf <http://zbw.eu/stw/version/8.08/delta/8.10> ;
  delta:deletion [] ;
  delta:insertion [ skos:altLabel "Employee selection"@en ] .

<http://zbw.eu/stw/descriptor/12571-4/version/8.10/delta/8.12>
  a sh:ConceptDelta ;
  dcterms:isPartOf <http://zbw.eu/stw/version/8.10/delta/8.12> ;
  delta:deletion [ skos:related <http://zbw.eu/stw/descriptor/11295-0> ] ;
  delta:insertion [] .

<http://zbw.eu/stw/descriptor/12571-4/version/8.04/delta/8.06>
  a sh:ConceptDelta ;
  dcterms:isPartOf <http://zbw.eu/stw/version/8.04/delta/8.06> ;
  delta:deletion [] ;
  delta:insertion [ skos:altLabel "Bewerbungsgespräch"@de ] .

<http://zbw.eu/stw/descriptor/12571-4>
  sh:conceptHistory <http://zbw.eu/stw/descriptor/12571-4/version/8.10/delta/8.12> , <ht
```

Future work and the skos-history project

- Apply to differing concept schemes
- Distill general properties useful for human-readable change reports as well as machine-actionable data

Please consider joining – particularly if

- you are in charge of a KOS and want to publish its change history
- you are using one or several KOS in an application, or intellectually, and want to trace and re-apply upstream changes
- just feel challenged by the task

Thanks for listening!

Joachim Neubert

ZBW – Leibniz Information Centre for Economics

j.neubert@zbw.eu

<http://zbw.eu/stw>

<https://github.com/jneubert/skos-history>

<http://zbw.eu/labs>