

KOS evolution in Linked Data

Joachim Neubert

ZBW – Leibniz Information Centre for Economics, Hamburg

SWIB14

Bonn, Germany

03.12.2014

Agenda

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

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 Data in SKOS
- Roughly every year a new version
- Major overhaul in progress – subject area by subject area

Short digression: SKOS as a RDF data format

- Based on *concepts* (“units of thought”), which may bear labels in multiple languages
- All semantic relations (hierarchies, mappings etc.) exist between concepts
- Per language at most one `skos:prefLabel` (should be unique)
- Additional properties for notations, notes, mappings, etc.
Classes for `ConceptSchemes` and `Collections` of concepts
- Widely in use today as a common interchange format

How did we handle KOS evolution in the past?

RDF statements about a particular version

```
<http://zbw.eu/stw>  
  a skos:ConceptScheme, void:Dataset ;  
  dcterms:issued "2013-10-30"^^xsd:date ;  
  owl:versionInfo "8.12" ;  
  ...
```

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>

303 redirect to versioned URLs (RDFa/rdf/ttl files)

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

Archived RDFa/rdf/ttl files available

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

(Currently, search functions and web services always work on the latest version)

Deprecated concepts

No deletion – 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 .
```


Pragmatic version history solution: Don't delete anything

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.

- [8.12 \(Detailed Changelog - in German\)](#)
- [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. Aserbaidshisch [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)

How to handle this better?

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 a derived subset of a KOS
- People maintaining mappings to other vocabularies, and applications supporting them
- Automatic or semi-automatic indexing applications which make use of the KOS and/or its mappings
- Search applications which make use of the KOS and/or its mappings

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.)

Dataset versioning + skos-history

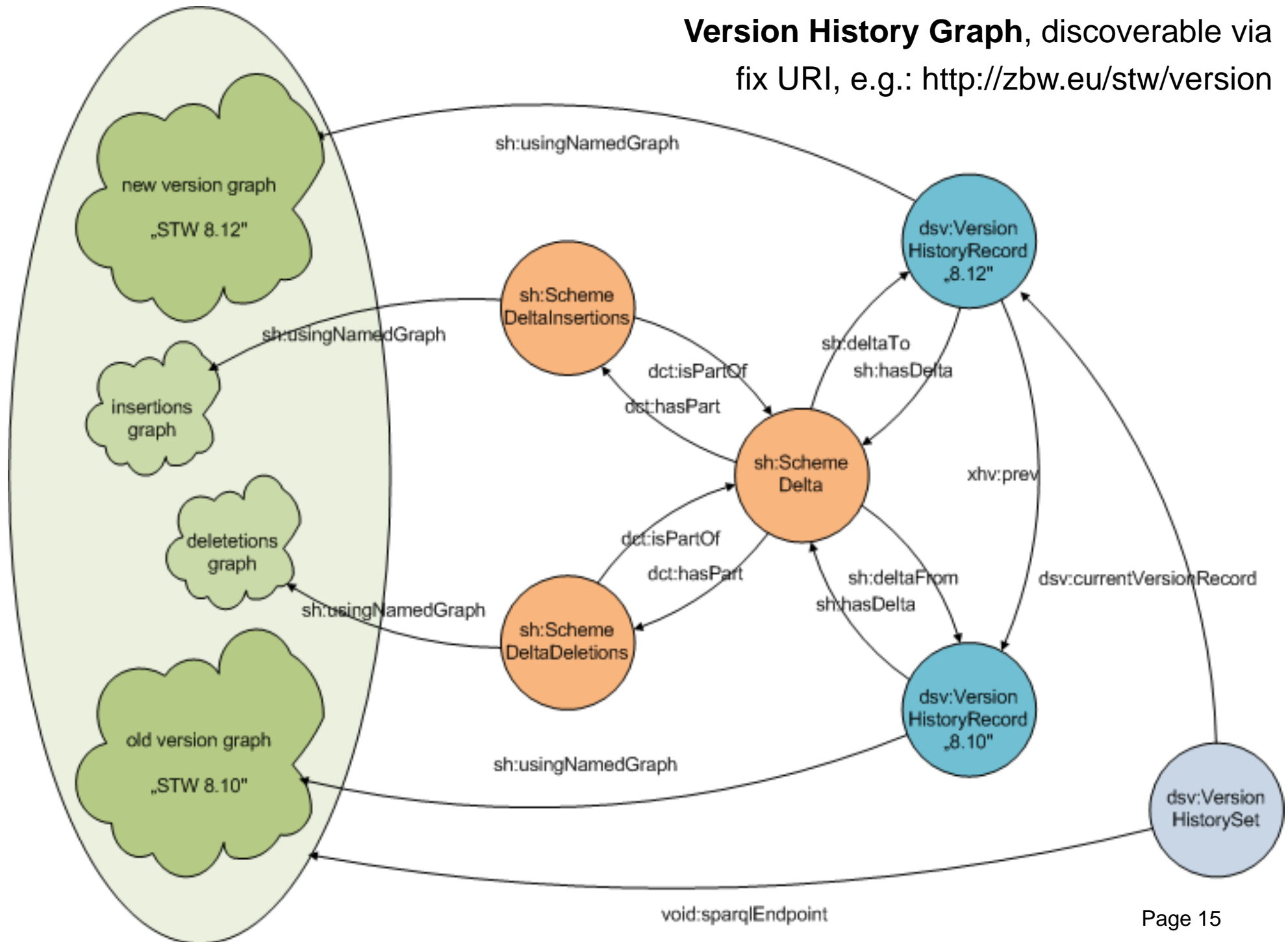
- should basically work on every SKOS vocabulary

5 basic steps to an actionable skos-history

- 1) Start with a sorted n-triple file per version.
(This poses one triple on every single line.)
- 2) Create a raw diff between two version files.
(This gives you thousands and thousands of differences, even excluding bnodes.)
- 3) Split the resulting diff into an insertions and a deletions file.
- 4) Load the version files, the insertions and deletions files into a triple store as named graphs.
- 5) Add metadata about the versions and the deltas in a separate „version history graph“.

https://github.com/jneubert/skos-history/blob/master/bin/load_versions.sh

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



Vocabularies 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](#)

New concepts by subject category

B.06 Production management

Cost efficiency

B.06 Production management

Job performance

B.06 Production management

Plant asset management

B.06 Production management

Production process

B.07 Marketing

Brand architecture

B.07 Marketing

Brand extension

B.07 Marketing

Brand loyalty

B.07 Marketing

Celebrity endorsement

B.07 Marketing

Consumer attitudes

B.07 Marketing

Consumer preferences

B.07 Marketing

Customer acquisition

B.07 Marketing

Customer integration

Statistics via aggregation queries: STW

Version	Date	Added descriptors	Deprecated descriptors	redirected	Added thsys	Deprecated thsys*
v 8.04	16.02.2009					
v 8.06	22.04.2010	224	4	4	3	
v 8.08	30.06.2011	131	57	54	14	1
v 8.10	21.03.2012	105	141	110	7	4
v 8.12	30.10.2013	260	487	485	12	26
v 8.14	18.11.2014	227	342	342	?	?

* Computed column - deprecation and redirects for thsys will be introduced for STW v 8.14 (retrospectively)

https://github.com/jneubert/skos-history/blob/master/bin/create_change_statistics.pl

Statistics via aggregation queries: TheSoz

Version	Date	Added concepts	Deleted concepts
v 0.7	11.01.2011		
v 0.86	08.11.2011	1	1
v 0.91	30.04.2012	240	4
v 0.92	19.09.2012	15	3
v 0.93	25.02.2014	42	4

Thesaurus for the Social Sciences

<http://www.gesis.org/en/services/research/thesauri-und-klassifikationen/social-science-thesaurus/>

https://github.com/jneubert/skos-history/blob/master/bin/create_change_statistics.pl

Selected useful reports

- Changed notations
- Splits and merges of concepts
- History of a single concept

Changed notations (general case)

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

Changed notations (linking STW versioned pages)

old	new
B.02.02.01 Capital Budgeting	B.02.03 Capital budgeting
B.06.02 Manufacturing Systems and Manufacturing Technology	B.06.03 Production organization
B.06.03 Computer-Aided Manufacturing	B.06.02 Factor input
B.11 Global Management	B.01.07 Global Management
N.04.04.01 Security Policy	N.04.04.05 Security Policy
N.04.04.02 International Conflicts	N.04.04.06 International Conflicts
N.04.04.04 European Integration and EU Policy	N.04.04.02 European integration
N.05.08 International Law	N.05.07 International law
N.05.08.01 Public International Law	N.05.07.01 Public international law
N.05.08.02 Community Law	N.05.07.02 Community law

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

- ▶ A General descriptors
- ▼ B Business economics
 - B.00 Business Economics
 - ▶ B.01 Management and Business Organization
 - ▼ B.02 Corporate Finance and Investment Policy
 - ▶ B.02.01 Corporate Finance
 - ▼ B.02.02 Corporate Investment Behaviour
 - B.02.02.01 Capital Budgeting
 - ▶ B.03 Business Accounting and Auditing
 - ▶ B.04 Human Resource Management
 - ▶ B.05 Materials Management and Logistics
 - ▶ B.06 Production Management
 - ▶ B.07 Marketing
 - ▶ B.08 Corporate Taxation and Accounting
 - ▶ B.09 Business Information Systems
 - B.10 Operations Research
 - B.11 Global Management
- ▶ G Geographic names
- ▶ N Related subject areas
- ▶ P Commodities
- ▶ V Economics
- ▶ W Economic sectors

B.02.02.01 Capital Budgeting

B.02.02.01 Investitionsplanung und -rechnung (german)

broader

- B.02.02 Corporate Investment Behaviour ▼

Descriptors

- Capital budgeting 
- Capital good 
- Corporate investment decision 
- Discounting 
- Disinvestment 
- Dynamic capital budgeting 
- Foreign direct investment 
- Investment 
- Investment risk 
- Investment theory by the firm 
- Mathematics of finance 
- Qualitative data analysis 
- Real option 
- Replacement investment 
- Sensitivity analysis 
- Useful life 

Persistent Identifier (for bookmarking and linking)

- <http://zbw.eu/stw/thesis/71044>

- ▶ [A General descriptors](#)
- ▼ [B Business economics](#)
 - [B.00 Business Economics](#)
 - ▶ [B.01 Management and Business Organization](#)
 - ▼ [B.02 Corporate finance and capital budgeting](#)
 - ▶ [B.02.01 Corporate finance](#)
 - [B.02.02 Corporate financial assets](#)
 - [B.02.03 Capital budgeting](#)
 - ▶ [B.03 Accounting](#)
 - ▶ [B.04 Human Resource Management](#)
 - ▶ [B.05 Materials Management and Logistics](#)
 - ▶ [B.06 Production management](#)
 - ▶ [B.07 Marketing](#)
 - ▶ [B.08 Corporate tax management](#)
 - ▶ [B.09 Business Information Systems](#)
 - [B.10 Operations Research](#)
- ▶ [G Geographic names](#)
- ▶ [N Related subject areas](#)
- ▶ [P Commodities](#)
- ▶ [V Economics](#)
- ▶ [W Economic sectors](#)



















B.02.03 Capital budgeting

B.02.03 Investitionsrechnung (german)

broader

- [B.02 Corporate finance and capital budgeting](#) ▼

Descriptors

- [Business mathematics](#) EB 
- [CAPM](#) EB 
- [Corporate investment theory](#) EB 
- [Decision under uncertainty](#) EB 
- [Disinvestment](#) EB 
- [Dynamic investment appraisal](#) EB 
- [Foreign investment](#) EB 
- [Greenfield investment](#) EB 
- [Investment](#) EB 
- [Investment appraisal techniques](#) EB 
- [Investment decision](#) EB 
- [Investment risk](#) EB 
- [Investment under uncertainty](#) EB 
- [Net present value method](#) EB 
- [Real options analysis](#) EB 
- [Reinvestment](#) EB 
- [Sensitivity analysis](#) EB 
- [Useful life](#) EB 

Persistent Identifier (for bookmarking and linking)

- <http://zbw.eu/stw/thesys/71044>

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

Concept removed and merged into multiple

Minor split-ups of concepts can be revealed by label movements, too:

Snack food

*Deprecated (used at last in version 8.12), USE
Convenience food*

deprecatedConcept	movedLabel	newConcept
Snack food	Kartoffelchips	Potato
Snack food	Käsegebäck	Bakery product
Snack food	Salzgebäck	Bakery product
Snack food	Schokoriegel	Chocolate

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

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 "Bewerbersauswahl"@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

- For STW:
 - Create a web service for concept history and link a history report to every concept
 - Provide drilldowns for new/deprecated/... concepts from the category level, perhaps visualizations / heat maps
- For skos-history:
 - Apply to differing concept schemes
 - Distill general properties useful for human-readable change reports as well as machine-actionable data

Consider joining the skos-history project ...

... 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

Code, issues, wiki pages etc.: <https://github.com/jneubert/skos-history>

Currently, Johan DeSmedt (Tenforce) , Sini Pessala (National Library of Finland) and Agis Papantoniou (Tenforce) are involved in the project and in discussions on which this presentation was based.

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>