

Change Tracking in Knowledge Organization Systems with skos-history

Joachim Neubert & Osma Suominen

ZBW – Leibniz Information Centre for Economics, Kiel/Hamburg &
The National Library of Finland, Helsinki

DCMI/ASIST/AIMS Webinar Series:

Generic Tools and Methods for SKOS-based Concept Schemes

16.3.2016

Agenda

- User questions and requirements
- Getting a grip on changes:
 - Overview
 - Creating a version store
 - Generic queries
 - Dataset-specific adaption of queries
- skos-history in use
 - Application at the National Library of Finland
 - Application for STW Thesaurus for Economics
- Outlook: Future work and the skos-history project

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 utilizing the KOS

Agenda

- User questions and requirements
- Getting a grip on changes:
 - **Overview**
 - Creating a version store
 - Generic queries
 - Dataset-specific adaption of queries
- skos-history in use
 - Application at the National Library of Finland
 - Application for STW Thesaurus for Economics
- Outlook: Future work and the skos-history project

Overview: 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 approach

=> should work on every SKOS vocabulary

Scope of vocabulary versioning

- Versioning the concept scheme, not each individual concept
- URIs for the concepts remain stable over the different versions
- Distinct versions of a vocabulary, or at least timestamped dumps, must be available
- Support for a continuous flow of changes, e.g., the LoC Subject Headings, or the concepts of the GND, is currently not provided

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“.

Agenda

- User questions and requirements
- Getting a grip on changes:
 - Overview
 - **Creating a version store**
 - Generic queries
 - Dataset-specific adaption of queries
- skos-history in use
 - Application at the National Library of Finland
 - Application for STW Thesaurus for Economics
- Outlook: Future work and the skos-history project

Hands on: Create a version store for skos-history

Requirements:

- SPARQL 1.1 compliant service or repository ('triple store'), accessible in read/write mode
<https://github.com/NatLibFi/Skosmos/wiki/InstallTutorial#install-jena-fuseki>
- An environment for executing bash scripts for the data load script (any Linux should do, Cygwin may).

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

Code of scripts and queries: also on [GitHub](#)

Load a version store: config file for JEL

```
root@ite-srv11:/opt/skos-history/bin
#?/bin/bash

# files are loaded from $BASEDIR/$VERSIONS/$FILENAME
VERSIONS=(20120320 20130621)
BASEDIR=/opt/thes/var/jelv
FILENAME=rdf/jel.ttl
SCHEMEURI="http://zbw.eu/beta/external_identifiers/jel"

# private read/write endpoints
PUT_URI=http://localhost:8080/fuseki/jelv/data
UPDATE_URI=http://localhost:8080/fuseki/jelv/update

# public read-only endpoint
QUERY_URI=http://zbw.eu/beta/jelv/sparql/query
~
~

"jel.config" 14L, 406C          14,1          All
```

Configuration for Fuseki (<https://github.com/jneubert/skos-history/blob/master/bin/jel.config>);
see also configuration for Sesame (<https://github.com/jneubert/skos-history/blob/master/bin/jel.sesame.config>)

Load a version store: load_versions.sh script

```
root@ite-srv11/opt/skos-history/bin
statement=""
$PREFIXES
insert {
  GRAPH <${delta_uri}/${op}> {
    ?s ?p ?o
  }
}
where {
  graph <${BASEURI}/${minuend}> {
    ?s ?p ?o
  }
  minus {
    graph <${BASEURI}/${subtrahend}> {
      ?s ?p ?o
    }
  }
}
# filter out blank nodes
filter isIRI(?s)
filter (isIRI(?o) || isLiteral(?o) || isNumeric(?o))
}
}

sparql_update "$statement"

"load_versions.sh" 418L, 10370C 255,3 61%
```

Load a version store: load_versions.sh script

```
root@ite-srv11:/opt/skos-history/bin
[root@ite-srv11 bin]#
[root@ite-srv11 bin]#
[root@ite-srv11 bin]#
[root@ite-srv11 bin]# ./load_versions.sh -f jel.config

Initializing the version history graph with the current version

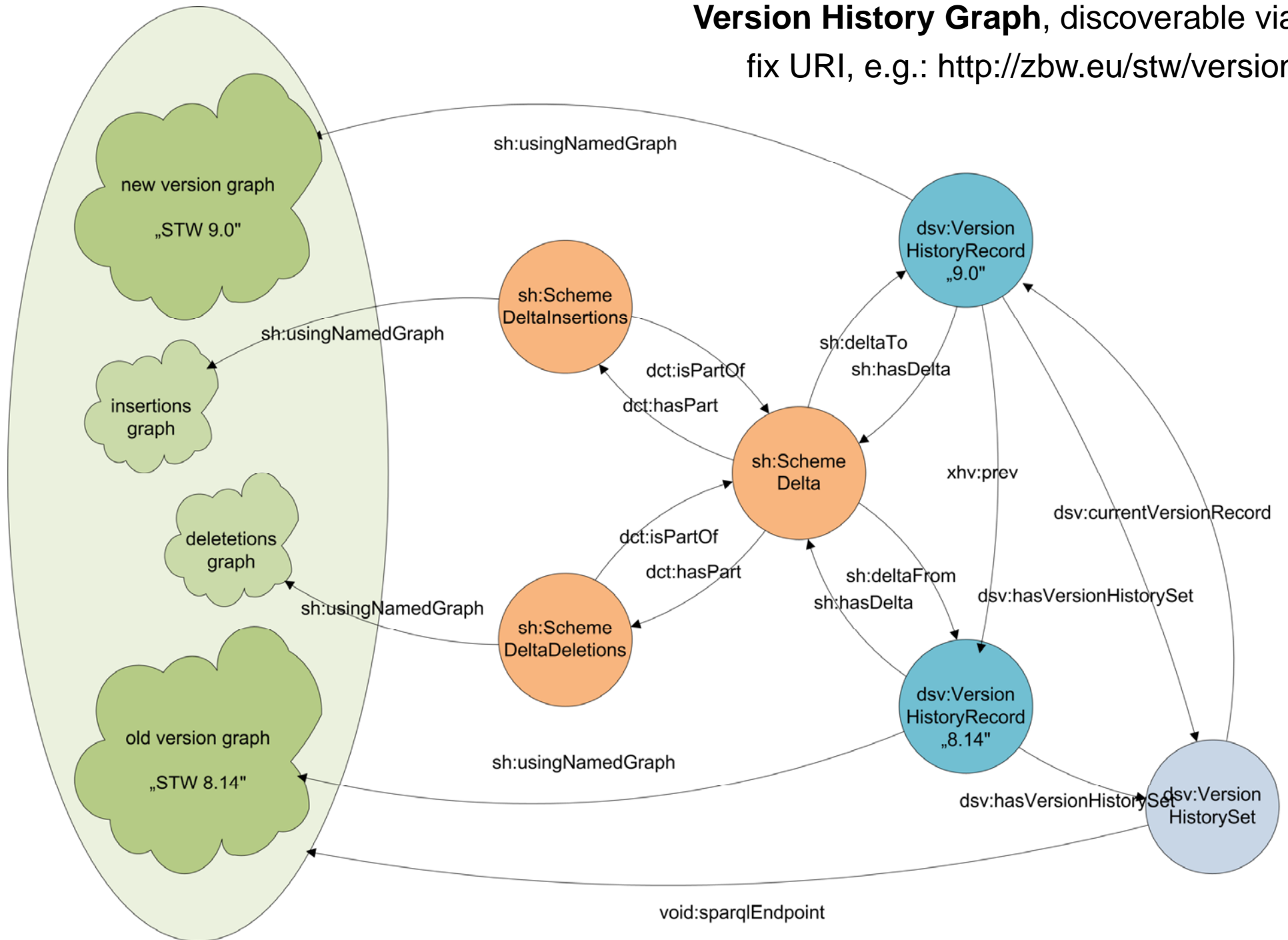
Loading version http://zbw.eu/beta/external_identifiers/jel/version/20120320

Loading version http://zbw.eu/beta/external_identifiers/jel/version/20130621

Creating the delta http://zbw.eu/beta/external_identifiers/jel/version/20120320/delta/20130621

[root@ite-srv11 bin]#
[root@ite-srv11 bin]# █
```

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



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

Version History Graph, published as HTML/RDFa

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)

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/version/latest/descriptor/19664-4/about.en.html>).

<http://zbw.eu/stw/version>

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!

Agenda

- User questions and requirements
- Getting a grip on changes:
 - Overview
 - Creating a version store
 - **Generic queries**
 - Dataset-specific adaption of queries
- skos-history in use
 - Application at the National Library of Finland
 - Application for STW Thesaurus for Economics
- Outlook: Future work and the skos-history project

```

SELECT distinct (?concept AS ?addedConcept) (str(?prefLabel) AS ?addedConceptLabel)
WHERE {
  GRAPH ?versionHistoryGraph {
    # parameters
    VALUES ( ?versionHistoryGraph ?oldVersion ?newVersion ?language ) {
      ( undef undef undef "en" )
    }
    # get the current and the previous version as default versions
    ?versionset dsv:currentVersionRecord/xhv:prev/dc:identifier ?previousVersion .
    ?versionset dsv:currentVersionRecord/dc:identifier ?latestVersion .
    # select the versions to actually use
    BIND(coalesce(?oldVersion, ?previousVersion) AS ?oldVersionSelected)
    BIND(coalesce(?newVersion, ?latestVersion) AS ?newVersionSelected)
    # get the delta and via that the relevant graphs
    ?delta a sh:SchemeDelta ;
      sh:deltaFrom/dc:identifier ?oldVersionSelected ;
      sh:deltaTo/dc:identifier ?newVersionSelected ;
      sh:deltaFrom/sh:usingNamedGraph/sd:name ?oldVersionGraph .
    ?insertions a sh:SchemeDeltaInsertions ;
      dcterms:isPartOf ?delta ;
      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

```

Newly inserted concepts – results

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](#)

Reports operating on standard SKOS structures

SPARQL queries

The queries linked from this page provide examples on how a skos-history version store can be utilized to create useful change reports. The queries can be executed via an [interactive SPARQL GUI](#) against an [example endpoint](#) with different versions of [STW Thesaurus for Economics](#), prepared as a skos-history version store according to [Versions and Deltas as Named Graphs](#).

Generic queries for any SKOS vocabulary

Overview

Query	Description
version_overview	Discover the version history graph and all available scheme versions

Lists of concepts

Query	Description
added_concepts	Identify all concepts inserted in the current version
added_concepts_with_top_concepts	Identify all concepts inserted in the current version with their top concepts
labels_moved_to_added_concepts	Show the labels which have moved to newly inserted concepts (flag new concepts if subordinated to the old one)
deprecated_concepts	Identify all concepts deprecated with the current version
deleted_concepts	Identify all concepts deleted with the current version
changed_notations	For a classification (in this case the subject categories of STW), show which notation has changed

Aggregated information about versions

Query	Description
count_added_concepts	Count concepts inserted per version

Reports ... (continued)

History of selected concepts

Query	Description
concept_deltas	All version deltas for the concept Personnel selection (Insert other example concept uris into the VALUES clause - suggestions in the comment.)
concept_history	Early alternative approach for the history of the concept Personnel selection (changes in pref/allLabels only)

Technical background information

Query	Description
version_graph	Complete version history graph
service_graph	Complete service description graph (default graph)

Extension to SKOS-XL

Some of the example queries have been adapted to work against version of thesauri, which uses [SKOS-XL](#). The queries can be directed to the [thesoz](#) or [agrovoc](#) example endpoints (see below).

Dataset-specific queries

More often than not, SKOS publications contain information specific to the dataset in question. Dataset-specific queries may exploit and expose this additional information.

- [STW Thesaurus for Economics](#) (beta)
- [GND subject headings](#) (experimental)

So while the examples will not run with other datasets, they may show how value can be added by exploiting custom data structures, while at the same time making use of a common version history scheme.

Further (purely experimental) skos-history example endpoints

- TheSoz - add `&endpoint=http://zbw.eu/beta/sparql/thesozv/query` to the URL
([version overview](#))
- YSO - add `&endpoint=http://zbw.eu/beta/sparql/ysov/query` to the URL
([version overview](#))

Changed notations

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

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

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

```

Agenda

- User questions and requirements
- Getting a grip on changes:
 - Overview
 - Creating a version store
 - Generic queries
 - **Dataset-specific adaption of queries**
- skos-history in use
 - Application at the National Library of Finland
 - Application for STW Thesaurus for Economics
- Outlook: Future work and the skos-history project

GND subjects by subject category – query

```

# for each inserted concept, a newly inserted prefLabel must exist ...
GRAPH ?insertionsGraph {
  ?concept skosxl:prefLabel/skosxl:literalForm | skos:prefLabel ?prefLabel ;
  gndo:gndSubjectCategory ?category
.
}
# ... and the concept must not exist in the old version
FILTER NOT EXISTS {
  GRAPH ?oldVersionGraph {
    ?concept ?p []
  }
}
GRAPH ?newVersionGraph {
  ?category skos:prefLabel ?catLabel ;
  skos:notation ?notation .
}
ORDER BY ?category ?prefLabel

```

GND subjects by subject category – results

Showing 1 to 25 of 25 entries (filtered from 736 total entries)

Search: Show entries

notation	category	addedConcept
1 30	Informatik, Datenverarbeitung	Cyberattacke
2 30	Informatik, Datenverarbeitung	Desktop-Computer
3 30	Informatik, Datenverarbeitung	Fortgeschrittene elektronische Unterschrift
4 30	Informatik, Datenverarbeitung	Industrie 4.0
5 30	Informatik, Datenverarbeitung	Machine to Machine
6 30	Informatik, Datenverarbeitung	Mikroservice
7 30	Informatik, Datenverarbeitung	Namensschild
8 30	Informatik, Datenverarbeitung	Online-Community
9 30	Informatik, Datenverarbeitung	Produktionstechnik
10 30	Informatik, Datenverarbeitung	Proxy <Entwurfsmuster>
11 30	Informatik, Datenverarbeitung	Serious game
12 30	Informatik, Datenverarbeitung	Soft Error
13 30	Informatik, Datenverarbeitung	Text-to-Speech

STW deprecated concepts – query

```
# identify the deprecated concepts
GRAPH ?insertionsGraph {
  ?concept owl:deprecated true
}
# get the replacedBy information
GRAPH ?newVersionGraph {
  # get concepts and restrict to conceptType
  ?concept a ?conceptType .
  OPTIONAL {
    ?concept dterms:isReplacedBy ?replacedBy .
    ?replacedBy skos:prefLabel ?replacedByPrefLabel .
  }
}
# get categories
```

STW deprecated concepts – result

Showing 1 to 27 of 27 entries

Search: Show entries

	secondLevelCategory	deprecatedConcept	replacedByConcept
1	G.02 Asia	Greater China	China
2	N.04 Politics and Political Science	Constituency	Electoral system
3	N.04 Politics and Political Science	Government system	Political system
4	N.04 Politics and Political Science	National symbol	State
5	N.04 Politics and Political Science	Occupation of territory	Military intervention
6	N.04 Politics and Political Science	Parliamentary allowance	Political finance
7	N.04 Politics and Political Science	Parliamentary committee	Parliament
8	N.04 Politics and Political Science	Parliamentary group	Political party
9	N.04 Politics and Political Science	Party congress	Party politics

Agenda

- User questions and requirements
- Getting a grip on changes:
 - Overview
 - Creating a version store
 - Generic queries
 - Dataset-specific adaption of queries
- skos-history in use
 - **Application at the National Library of Finland**
 - Application for STW Thesaurus for Economics
- Outlook: Future work and the skos-history project

skos-history at the National Library of Finland

see separate slides at <http://tinyurl.com/skos-history-nlf>

Agenda

- User questions and requirements
- Getting a grip on changes:
 - Overview
 - Creating a version store
 - Generic queries
 - Dataset-specific adaption of queries
- skos-history in use
 - Application at the National Library of Finland
 - **Application for STW Thesaurus for Economics**
- Outlook: Future work and the skos-history project

STW Thesaurus for Economics

- created in the 1990s
- on the web and available as SKOS since 2009
- bilingual (German/English)
- about 6000 descriptors, 500 subject categories
- overhaul during the last five years (five consecutive versions)

STW change reports (precompiled query results)

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

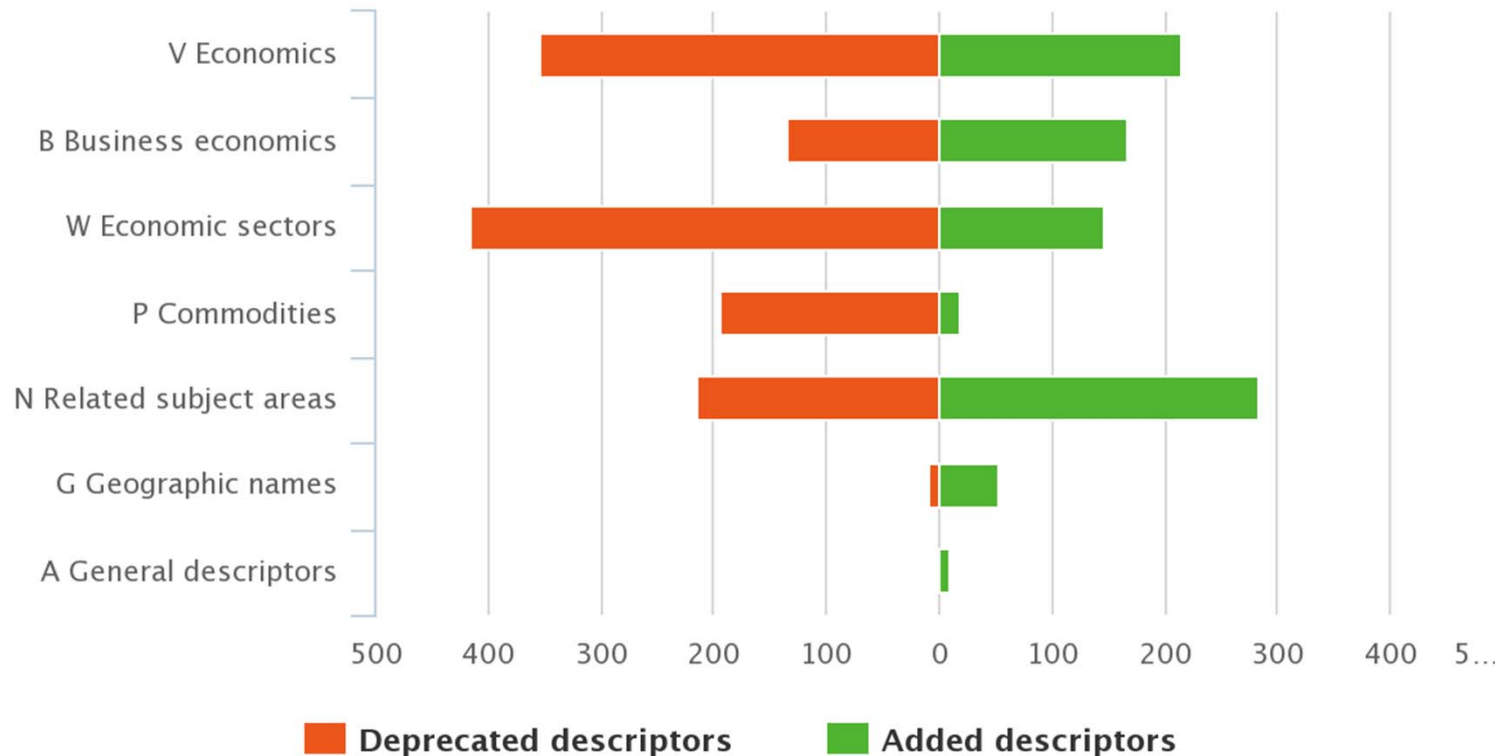
Subject categories:

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

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

Live Demo



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

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
- Get a grip on clusters of interrelated changes

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

Osma Suominen

The National Library of Finland

osma.suominen@helsinki.fi

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