



Case Study: Publishing STW Thesaurus for Economics as Linked Open Data

Timo Borst and Joachim Neubert, German National Library of Economics (ZBW), Germany

June 2009



General Description

The ZBW German National Library of Economics—Leibniz Information Centre for Economics is the world's largest economics library. It holds more than four million media items such as books, articles, journals, grey literature and databases. ZBW supports its users with fine-grained thematic access to these information resources. For this purpose the [STW Thesaurus for Economics](#) has been developed and applied since the 1990s. It provides a high-level taxonomy of subject categories, thousands of keywords (“descriptors”) and tens of thousands of both synonyms and links between the thesaurus concepts. The media items are indexed with descriptors from this thesaurus. They can be retrieved by these descriptors through the library catalog ECONIS.

The challenge

Although the STW has become an important in-house means for categorizing and cataloguing economics literature, there was still potential for increasing both access and reuse by other people and institutions. More concrete, we identified five action points: First, to improve web-based presentation of STW. Second, to foster precision of search results by actively suggesting preferred terms from STW. Third, to support the integration of STW into other indexing or retrieval environments. Fourth, to induce third-party reuse of the STW data, e.g. for customizing the vocabulary. And finally, to establish anchor points for linking to other vocabularies and datasets.

The solution

In order to provide a standard format for publishing the thesaurus as a whole, and also to decouple the publication process from the highly proprietary thesaurus maintenance application, we looked for a standardized, highly expressive intermediate format. It turned out that no common serialization format for thesauri yet exists. The “SKOS - Simple Knowledge Organization System” however, built within the Semantic Web community by vocabulary experts and targeting thesauri, classifications, folksonomies and the like, is expected to achieve the status of a W3C recommendation in 2009 and already has some implementations.

The mapping of the thesaurus concepts and relations to SKOS proved to be quite straightforward. Since SKOS is inherently multi-lingual, preferred and alternate labels (synonyms) in English could be attached to concepts as easily as their German equivalents. “Related”, “narrower” and “broader” relations mapped nicely to the according SKOS properties. Additional properties, such as publisher, version and licensing information were added seamlessly through the use of other RDF vocabularies (e.g., Dublin Core).

However, the crucial test for our SKOS adoption was the mapping of STW characteristics which are not very common among thesauri. In particular, STW has about 500 subject categories organizing about 6,000 descriptors in a single mono-hierarchic taxonomy. This taxonomy is not used for indexing, but as an aid for users to find appropriate indexing or search terms by browsing. We wanted to publish this taxonomy together with the descriptors in a single concept scheme. On the other hand, we still needed to distinguish them for rendering or for custom integrity checks. For these purposes SKOS provides extension points, and subclassing of SKOS concepts proved to meet our requirements.

The screenshot shows a web interface for the STW Thesaurus. On the left is a navigation menu with a breadcrumb trail: Home > Alphabetical descriptor list. The menu items are: A General descriptors, B Business economics, G Geographic names, N Related subject areas, P Commodities, and V Economics. Under V Economics, there are sub-items: V.00 Economics, V.01 Economic Theory, and V.02 Microeconomics. On the right, the selected category 'V.05.06 Financial Markets' is displayed in a larger font. Below it, the German name 'V.05.06 Finanzmärkte (german)' is shown. The interface also displays 'broader' and 'narrower' relationships. Under 'broader', there is a link to 'V.05 Money and Financial Markets'. Under 'narrower', there are four links: 'V.05.06.01 Money Market', 'V.05.06.02 Securities Markets', 'V.05.06.03 Forward Market', and 'V.05.06.04 Foreign Exchange Market'.

- Home
- Alphabetical descriptor list

- ▶ A General descriptors
- ▶ B Business economics
- ▶ G Geographic names
- ▶ N Related subject areas
- ▶ P Commodities
- ▼ V Economics
 - V.00 Economics
 - ▶ V.01 Economic Theory
 - ▶ V.02 Microeconomics
 - ▶ V.03 Macroeconomics
 - ▶ V.04 Business Cycle, Growth, and Economic Structure
 - ▼ V.05 Money and Financial Markets
 - V.05.01 Monetary Theory
 - V.05.02 Monetary Policy
 - V.05.03 Inflation
 - V.05.04 Interest Rate
 - V.05.05 Credit
 - ▼ V.05.06 Financial Markets
 - V.05.06.01 Money Market
 - ▶ V.05.06.02 Securities Markets
 - V.05.06.03 Forward Market

V.05.06 Financial Markets

V.05.06 Finanzmärkte (german)

broader

- V.05 Money and Financial Markets

narrower

- V.05.06.01 Money Market
- V.05.06.02 Securities Markets
- V.05.06.03 Forward Market
- V.05.06.04 Foreign Exchange Market

Descriptors

- Arbitrage
- Arbitrage pricing
- Beta Estimate
- Bid-ask spread
- Bubbles
- Capital asset pricing model
- Commodity exchange
- Credit market
- Effective rate
- Efficient-market hypothesis
- Euromarket
- Exchange
- Exchange listing
- Financial crisis
- Financial derivative

Figure 1: The STW taxonomy forms a high level knowledge organization system of economics and business management

Publishing the data on the Web was one of the main goals of the project. From the RDF file, we generated an XHTML page for each concept in the thesaurus, and embedded all of the data into this page using RDFa. We assigned a persistent, language- and version-independent URI to each page. Thus, the set of pages forms a highly interlinked network of semantic relations, usable for both humans and machines. Web server content negotiation is used to deliver the format (RDF/XML or XHTML, English or German) most appropriate to the request.

Financial crisis 

Finanzmarktkrise (german)

used for: Financial instability

Narrower Terms

- Banking crisis 
- Currency crisis 
- Stock market crash 

Broader Terms

- Economic crisis 

Related Terms

- International financial market 

Subject Categories

- V.05.06 Financial Markets
- V.07.07 International Financial Markets

Persistent Identifier (for bookmarking and linking)

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

Figure 2: An STW descriptor with its relations to other concepts and retrieval links into the library catalog

Providing links to other resources inside and outside the library was another main target of the new STW web presentation. The ZBW's own library catalog ECONIS was a natural choice for this. Users of the STW website can browse the pages and trigger a search for indexed media items by clicking on the book icon. This opens thematic access paths for the retrieval of library resources. Additional links were created to dbpedia entities, which enable linking to Wikipedia pages.

Since the development of the STW was publicly funded, we felt it was our obligation to publish it for free reuse. Under a non-commercial Creative Commons License the STW is now part of the Web of open linked data also in legal terms.

Future Work

ZBW plans to provide interfaces and widgets to select and aggregate descriptors for retrieval (eg. to provide compound search expressions like "financial crisis" OR "banking crisis" OR "stock market crash"). For searches in fields with controlled values, inclusion of narrower terms may lead to a more complete result set. For full text searches, the addition of synonyms and the addition of English or German forms may enhance searches. Other widgets may be used for indexing, e.g. to support authors in tagging their uploaded papers on a document server.

We also will create or make use of existing mappings to other vocabularies in economics, e.g. the classification system of the "Journal for Economics Literature" (JEL), and put them on to the Web as linked data.

As part of a future linked data infrastructure, ZBW has already deployed an experimental SPARQL endpoint (based on Joseki server) and a [terminology web service](#) built on the endpoint. The service provides, for example, search for concepts, narrower terms or synonyms. It also powers the autosuggest incremental search service on the STW web site.



Figure 3: Autosuggest search based on thesaurus web service

Key Benefits of Semantic Web Technology

- SKOS provided a quasi-standard serialization format for terminologies, supporting multilingual vocabularies and offering extensions for custom requirements
- additional information (version, license etc.) can be added seamlessly using other vocabularies
- RDFa easily allows enrichment of web pages containing thesaurus concepts with semantic annotations
- together, these technologies define URIs for concepts as reference points on the web and make third party reuse possible
- tools (e.g., triple stores and SPARQL servers) provide uniform, networked data access
- for the ZBW and other libraries, STW publication as linked open data opens up a broad range of more extensive applications, especially for data integration from different sources and search support
- regarding the field of economics and its resources, the STW provides an interlinking hub and an initial entry point into the web of the linked data