

Fusing Corporate Thesaurus Management with

Linked Data using PoolParty

Thomas Schandl









### PoolParty at a glance

- Developed by punkt. netServices
   Current release: PoolParty 2.8
- Main focus on three application areas:
  - SKOS Thesaurus Management
  - Linked Data (publishing & consuming)
  - Semantic Search & Semantic Indexing







## Challenge for Content Management



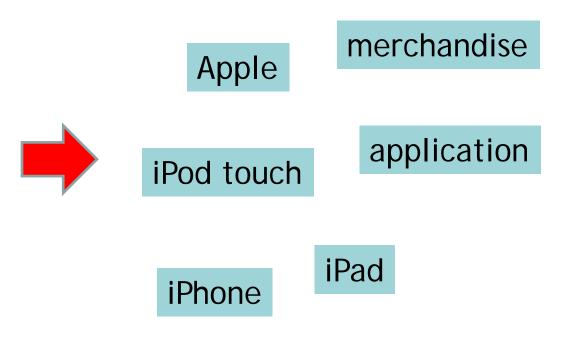
- 1. Annotation: Add meaning to the content
- 2. Link content: Bring content together in a meaningful way
- 3. Make content searchable: Add background knowledge to the content





## Traditional approach to annotate content with metadata

Apple is in the process of launching an application to allow iPhone, iPad and iPod Touch users to purchase Apple merchandise straight from their devices.





Semantic Web approach: Concepts & Relations instead of simple text

Apple http://my.com/Apple Apple is in the process of launching an Apple Inc. application to allow iPhone, iPad and iPod Touch users to purchase http://my.com/smartphone Apple merchandise straight from their http://my.com/iPhone devices. **iPhone** iPhone 3G http://my.com/iPhone3G iPhone 3GS





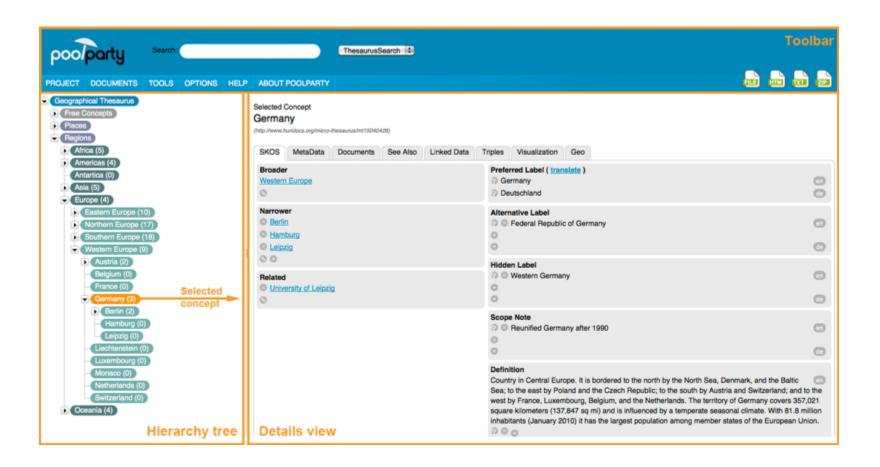
# poolporty in a nutshell

- W3C Semantic Web standards: Management of multi-lingual (corporate) thesauri & taxonomies on top of Semantic Web standards (SKOS, RDF, OWL & SPARQL)
- Usability: easy-to-use, web-based AJAX user interface
- Scalable Semantic Technologies: RDF Triple Store (SAIL), (Lucene) index engine and a phrase-extraction component
- Service oriented: PoolParty Server offers a Java-API & several interfaces: HTTP web services, SPARQL endpoint, Linked Data





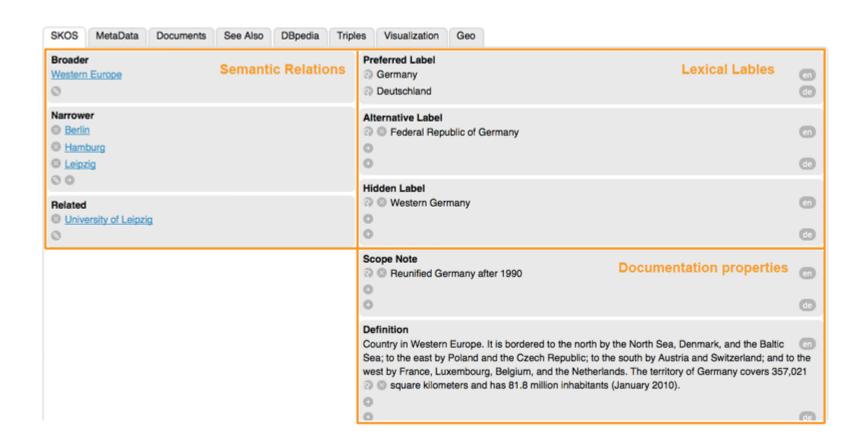
### **PoolParty GUI**







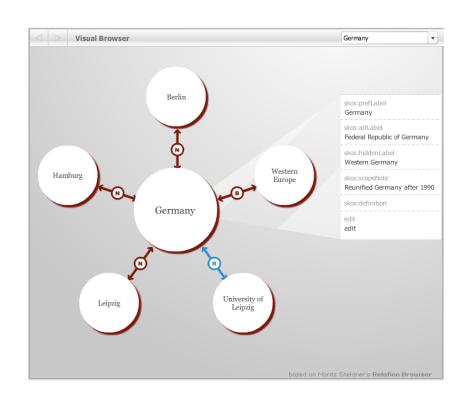
### Full compatibility with SKOS/RDF





### Some highlights: PoolParty thesaurus management

- Drag & drop , Auto-Complete
- Document analysis: phrase extraction
- Enrich concepts by using linked data
- Publish thesauri as linked data
- Advanced reporting functionality
- Import and validation of thesauri and CSV files
- Thesauris quality checker
- Wiki style collaborative editing of thesauri
- Visual browsing and map navigation





### Built-in automatic phrase extraction

- Supports different formats (html, doc, pdf, ppt, ...)
- Thesaurus based extraction
- Integrable with CMS, CRM etc.

#### Document Text

The city of Berlin, Germany is built around the River Spree in Bundesland, the federal state of Brandenburg. Berlin is Germany's capital city and has undergone enormous changes in the past 100 years. No longer divided, this city is now a thriving, modern and exciting destination, welcoming tourists and business visitors alike. Since World War II, the city has emerged from the past and undergone the biggest construction project in Europe. Much of Berlin has been rebuilt, restored or rejuvenated and there is some wonderful architecture, both old and new, standing side by side. Many come here for the exceptional nightlife alone, while others prefer the tourism aspect offered by the city. Those who choose to sightsee by day will discover that many of Berlin's landmarks come with their very own observation decks, such as the Reichstag (the German Parliament Building), the Europa Center on the Zoologischer Garten, the Kollhoff Tower on the Potsdamer Platz, the Siegessäule (Victory Column) in the Tiergarten area, and the Fernsehturm on the Alexanderplatz, the latter of which is a television tower and the tallest structure in the city, boasting a revolving restaurant. Popular districts with appealing accommodation include Charlottenburg, Friedrichshain, Kreuzberg, Mitte, Prenzlauer Berg and Wilmersdorf. When the Berlin Wall was destroyed is 1989 this was an enormous turning point for the city, which has previously been divided for over 28 years. East and West Berlin are now thoroughly unified, although there are still some signs of division in these areas. There are countless landmarks throughout the city and popular sights include the Berliner Dom - a huge domed cathedral standing within the Mitte borough; the Brandenburg Gate - the former city gate and now an iconic symbol of Germany; the Potsdamer Platz - one of the busiest squares in Europe, full of shops, bars, cafes and places of entertainment, Checkpoint Charlie - probably the most famous part of the old Berlin Wall; and the imposing Reichstag, home to the German parliament and just as impressive today as when it was completed in 1884.

Alexanderplatz capital city Berlin Berlin Wall Brandenburg capital Europe divided enormous German Germany's include landmarks Parliament Sights Potsdamer Potsdamer Platz Reichstag Towns undergone Wall West Berlin vears

▼ Highlight Tags in Document Hide Unselected Tags

Add Tag

ave Tag Event





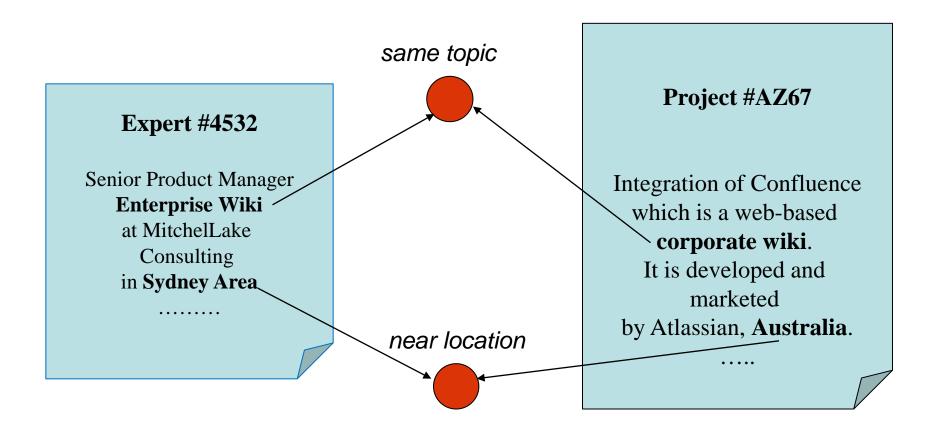
## Some Applications on top of PoolParty

- Tag recommendation: support users and content managers when annotating text
- Semantic Indexing: PoolParty TagEvent Store as a basis for a semantic index (→ IndexBuilder)
- Similarity search: "Similarity" is configurable: Certain features of a document can be "boosted" (example: persons, places / user tags etc.)
- Semantic Search and Navigation: Thesaurus can be used for facetted and moderated search (examples: emteba.at, ecoi.net)
- Search Engine Dictionaries: provide company or domain specific terms for search engine dictionary





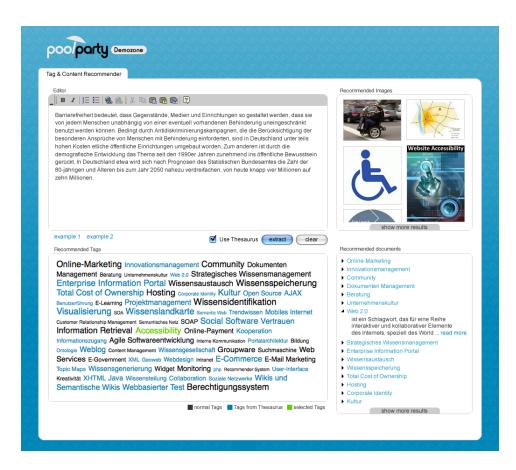
## Similarity search: finding the unexpected...







### PoolParty DemoZone



- compare thesaurus based approach with traditional approach
- tag recommender
- similar documents
- find images which fit to your document
- browser bookmarklet





### Wordpress Glossary Plugin

### GLOSSARY

This glossary will help to understand this blog much better. It was set up and is maintained with PoolParty. PoolParty glossaries can be integrated with the WordPress Blog-System easily. A wiki version of the glossary can be found here.

 $ALL \mid A \mid B \mid C \mid D \mid E \mid F \mid G \mid H \mid I \mid J \mid K \mid L \mid M \mid N \mid O \mid P \mid Q \mid R \mid S \mid T \mid U \mid V \mid W \mid X \mid Y \mid Z$ 

#### SEARCH ENGINE

(A full list of all glossary entries can be found here.)

A search engine is an information retrieval system designed to help find information stored on a computer system. The search results are usually presented in a list and are commonly called hits. Search engines help to minimize the time required to find information and the amount of information which must be consulted, akin to other techniques for managing information overload. The most public, visible form of a search engine is a Web search engine which searches for information on the World Wide Web. (Source: Wikipedia ©)

Broader: enterprise 2.0 platform

Narrower: index engine, search interface, crawler

Related: search

Search for search engine

 automatic generation of glossaries for Wordpress blogs

- SKOS compatibility
- automatic link detection and linkage with glossary term





### Programmatic access via Web Services

- getProposedTagsForDocument
- addTaggingEvent
- getTagFrequencies
- addDocumentToSimilarityIndex
- findSimilarDocuments
- getConceptSuggestions
- •



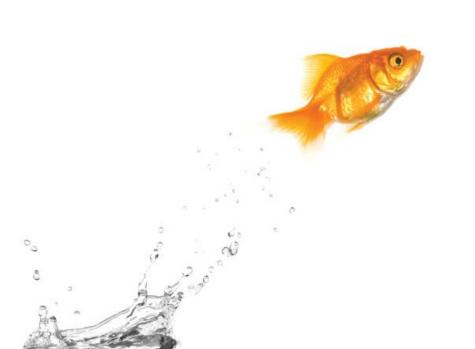


# Programmatic access - Example: emteba.at





Linked Data Features in Detail

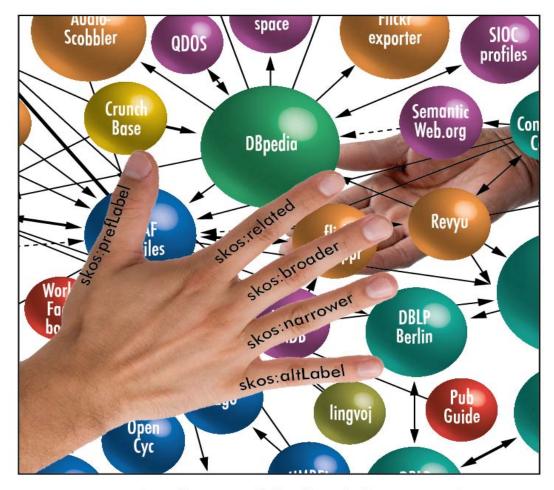








### SKOS Thesauri + Linked Data



Giving and taking of linked data with SKOS





### Linked Data - Benefits & Application Scenarios



### **Thesaurus Management**

- Automatic population of thesauri
- (Semi) Automatic categorization of new concepts

### **End User**



- Content augmentation
- Improved recommender services
- Improved navigation elements, e.g. in webshops

### **Content Provider**

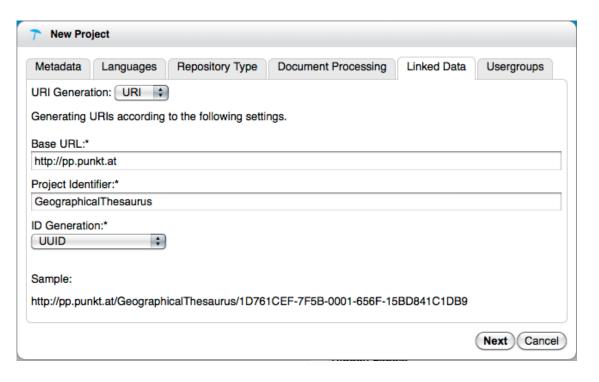


- Improved SEO
- Reduced costs of content management
- New services and mashups





## Publishing Linked Data with PoolParty



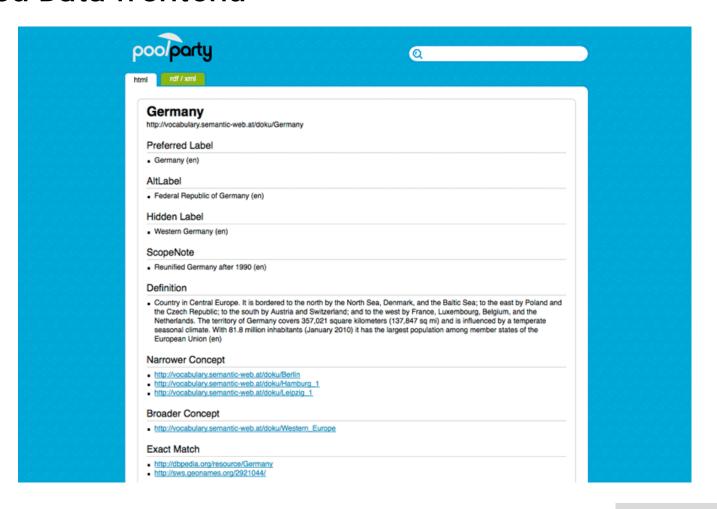
- using linked data patterns and "Cool URIs"
- Linked Data frontend

Additionally:

- Wiki front-end
- SPARQL-endpoint

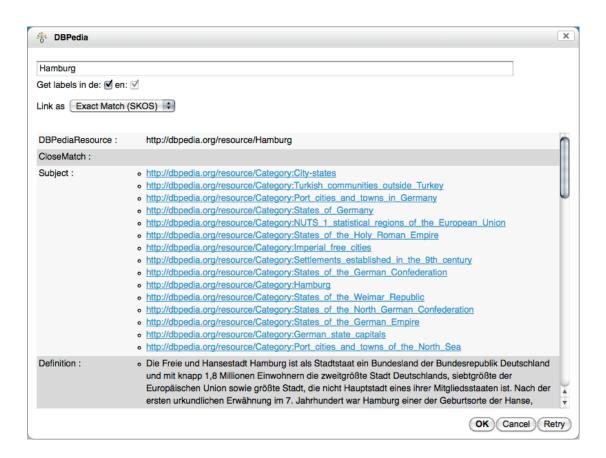


### Linked Data frontend





### Consuming Linked Data



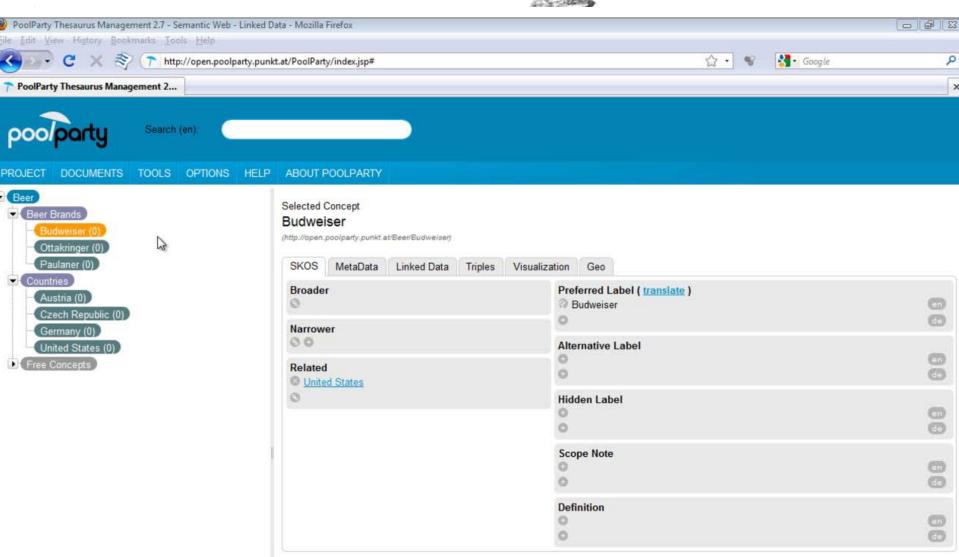
- advanced linked data look-up services
- expandable number of linked data sources already integrated
- linked data synchronisation mechanisms (beta)





### aus Ideen werden Erfolge

punkt. die webExperten







# Using SKOS context to link concepts to LD resources and semi-automatic population of thesaurus

Example: Thesaurus about arts and artists
Concept "Painters" with NT:
Kandinsky, Rembrandt and Berners-Lee

- Using broader and sibling concepts to help disambiguate and suggest the painter Berners-Lee
- Finding mutual categories from Dbpedia or Freebase
- Suggesting more NTs for Painters using LD categories



Semantic Search









# More background knowledge from thesauri and linked data can improve semantic search



- better disambiguation of search terms
- background knowledge of search terms help to "expand queries"
- better similarity search because of more metadata
- content augmentation through linked data





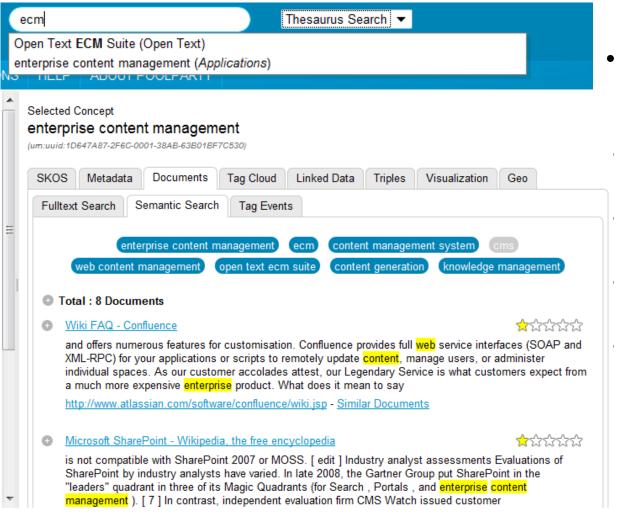
# Semantic Services provided by PoolParty

1	Search assistants (Auto-Complete, faceted search)		Improve user's search experience
2	Moderated Search	A PROPERTY OF THE PROPERTY OF	Creating complex queries
3	Tag Recommendation	Measurement Risk CSR firm Company lirr corporate performance business responsibility Criticism social responsibility good social firm with superior performance firm with poor reputation	Identifying the meaning of a document
4	Similarity Search (Recommender Systems)		Understanding relations





### Search Assistants



clever autocompletequery expansion

1 9 1

faceted search

visual search

Google synonyms





### **Moderated Search**

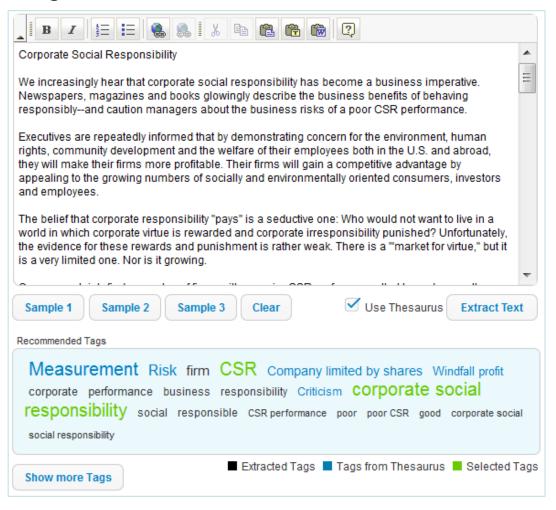
You searched in all countries for pentecostalism OR Protestantism in documents of any type published by any source
Refine your search with broader terms
(BT), narrower terms (NT), or related
terms (RT)
Protestantism
AND OR NOT
Christianity (BT)
Calvinism (NT)
Adventism (NT)
Pastors (NT)
Protestant Churches (NT)
show more
pentecostalism
AND OR NOT
Protestantism (BT)
Religious communities (RT)
SEARCH
SEARCH

- thesaurus helps to create complex queries
- supports multilinguality
- helps to explore a domain without deep knowledge





### Tag Recommendation

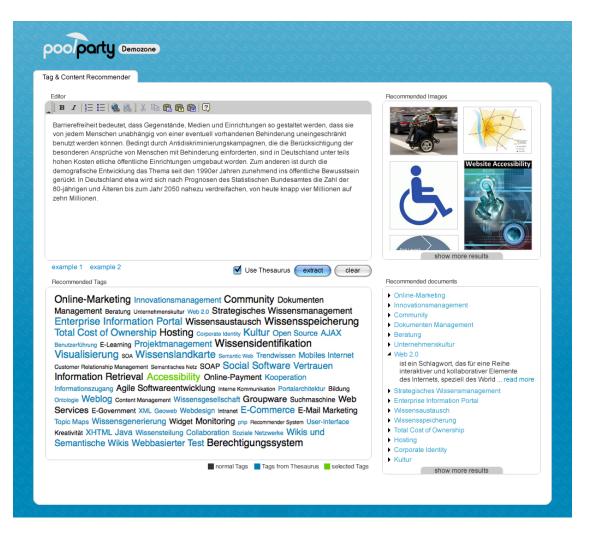


- annotation of documents with low effort
- motivation for people to annotate documents
- basis for building a semantic index





### Similarity Search



- improved similarity detection on top of additional background knowledge
- build recommender systems for web-shops or knowledge management systems
- help people to skim large document collections
- detect hidden relations between documents





## Integration of thesauri with Enterprise Search

Possible integrations with enterprise search engine:

- SC PD: TI KIS
- Autocomplete-Server
- Entity dictionary
- Query rewriting
- Moderated search
- Enrich semantic index

PoolParty Web-Services Integrate thesauri into search engine with real-time queries



PoolParty Reporting
Export parts of
thesauri into
individual XMLformats and
synchronize with
search engine

- improved semantic enterprise search
- all metadata can be administrated at one single place
- expandable via linked data mechanisms



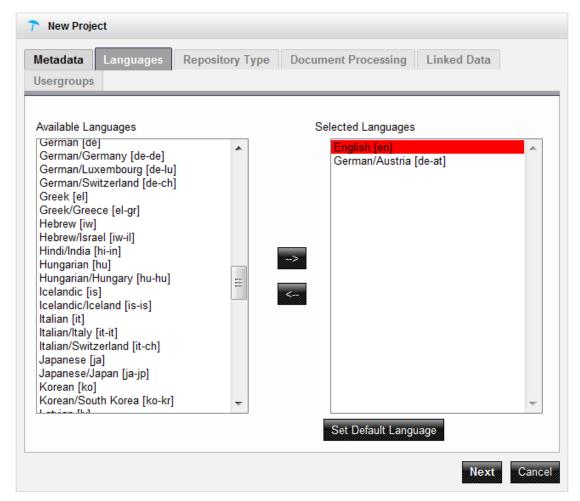
Thesaurus Management Advanced Features







### Multilinguality

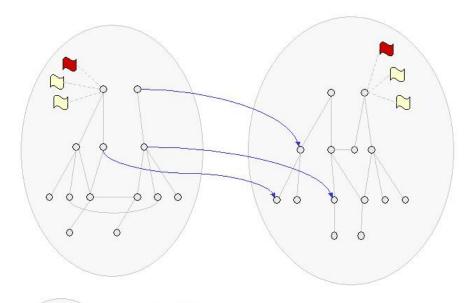








### Concept mapping



Concept scheme

Concept

Semantic relation

Semantic mapping

Preferred label

Alternative label

skos:exactMatch

skos:closeMatch

- → used for linked data mapping
- → used for concept mapping, e.g. after having imported a thesaurus





### Associating notes with concepts

KOS N	letadata	Linked Data	Triples	Visualization	Geo	
Notes I	History					
Note Type:	All Note	s	•			
Author:	All auth	ors	•			
From/To:						Show Notes
Concept H	istory Note					
Concept H	istory Note	past state/use	/meaning of	a concept.		
Concept H	istory Note about the	past state/use	/meaning of	a concept.		
Concept H Information  2010-08 Concept E	story Note about the 3-02 13:58 ditorial No	e past state/use :50 - Editorial te	/meaning of Note ( nagy	a concept.		
Concept H Information  2010-08 Concept E Information	about the a-02 13:58 ditorial No a for an ed 3-02 13:57	past state/use :50 - Editorial te itor, translator of :44 - Change N	/meaning of Note ( nagy or maintainer	a concept.  hel )  of the concept.		

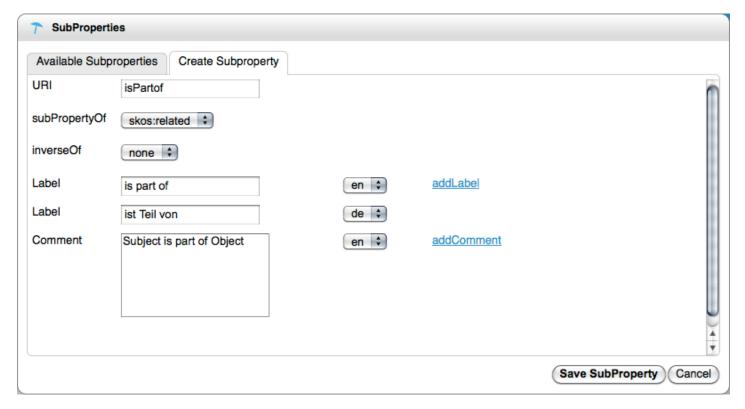
- skos:historyNote
- skos:changeNote
- skos:editorialNote

- used to trace meanings of a concept
- → used to discuss meanings of a concept





# Introduce individual relations between concepts

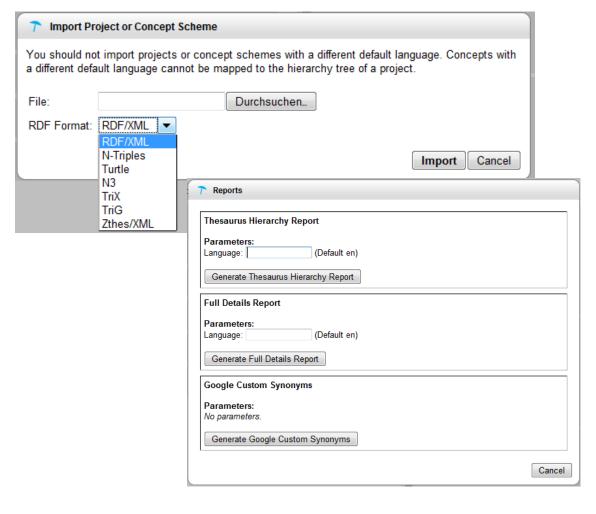


Create your own individual inverse or symmetric relations between concepts





### Import / export / reporting

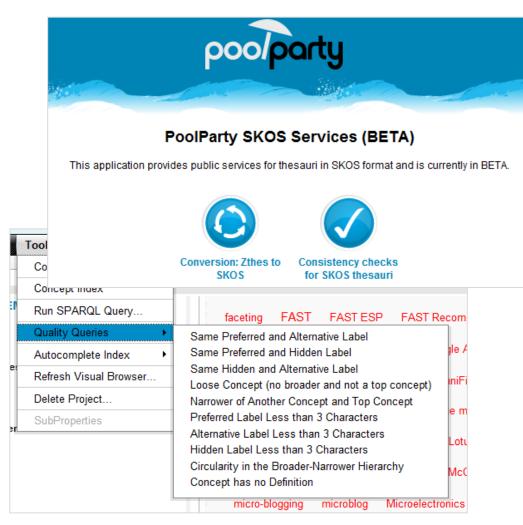


- import & export of SKOS using various RDF serializations
- import of CSV
- import of Zthes
- import/export of subtrees
- custom reports and XML exports based on PoolParty´s template engine





# Quality checks and validation service



Check thesauri to....

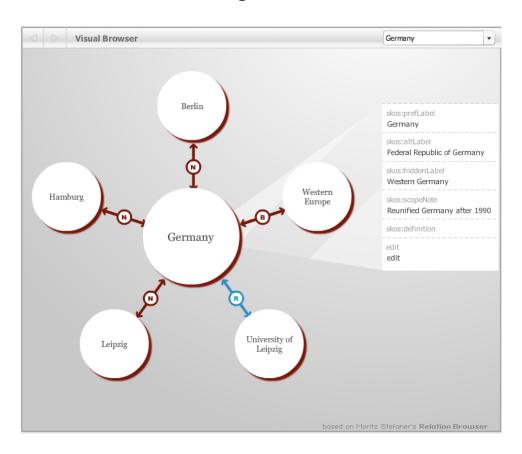
- be complete
- be non-cyclic (e.g. no circularity in the broader/narrower hierarchy).
- have no disjoints between related and hierarchical paths.

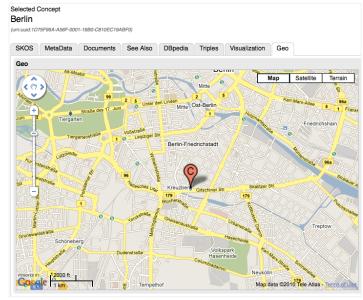




# aus Ideen werden Erfolge punkt. die webExperten

### Visual browsing





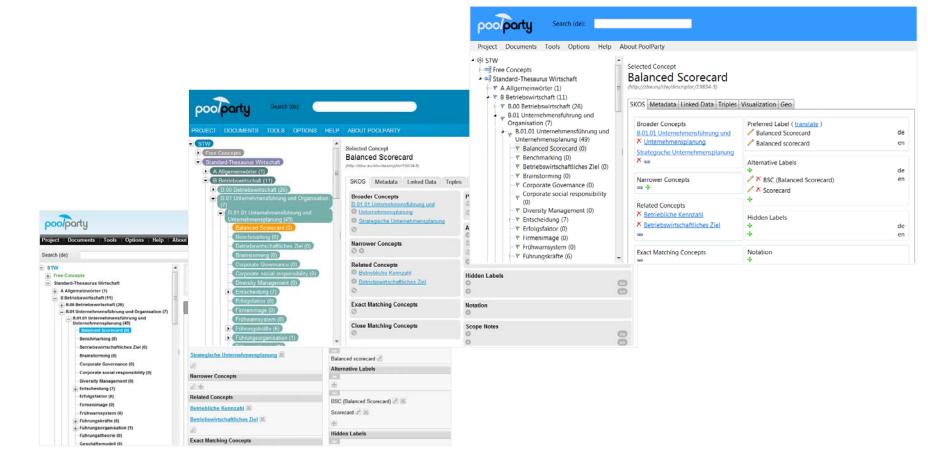






# aus Ideen werden Erfolge punkt. die webExperten

# Use your favourite theme!







### Contact

# Apply for a PoolParty demo account

http://poolparty.punkt.at/

### **Thomas Schandl**

schandl@punkt.at +43-1-8974122-27

### punkt. netServices GmbH

Lerchenfelder Guertel 43

A-1160 Wien / Austria

http://www.punkt.at/