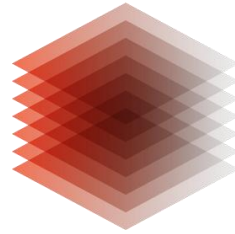


LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Integrating Data From Distributed Sources Via Lookup Services

Tatiana Walther <http://orcid.org/0000-0001-8127-2988>

Martin Barber <http://orcid.org/0000-0001-7924-0741>

Hamburg, 6.12.2017, SWIB2017



CC-BY 4.0 International
<https://creativecommons.org/licenses/by/4.0>

The Agenda of This Talk

1. Key Aspects
2. Lookup Services in VIVO – Current State
3. Motivations and Goals
4. Approaches
5. Skosmos
6. Technical Implementations
7. Demos: Destatis Fächersystematik, STW
8. Challenges & Future Plans

Integrating Data From Distributed Sources: Key Aspects

- **Purposes:** subject cataloguing, annotations, enrichment with further information through linking to referred objects etc.
- **Types of data:** subjects from thesauri and controlled vocabularies; things – persons, organizations, events, places
- **Formats:** RDF, SKOS, XML, non-machine-readable formats
- **Feasible ways of integration:** ingesting data dumps vs. direct access to external sources via lookup services



- Community-supported, open source software for representing scholarly activities
- Rests upon on Linked Data technologies
- Provides several external vocabulary sources

<https://wiki.duraspace.org/display/VIVODOC20x/Linking+to+External+Vocabularies>

Lookup Services in VIVO – Current State

Manage Concepts

There are currently no concepts specified.

External Vocabulary Services

- ☐ **AGROVOC** (Agricultural Vocabulary)
- ☐ **GEMET** (GEneral Multilingual Environmental Thesaurus)
- ☒ **LCSH** (Library of Congress Subject Headings)
- ☐ **UMLS** (Unified Medical Language System)

REST API

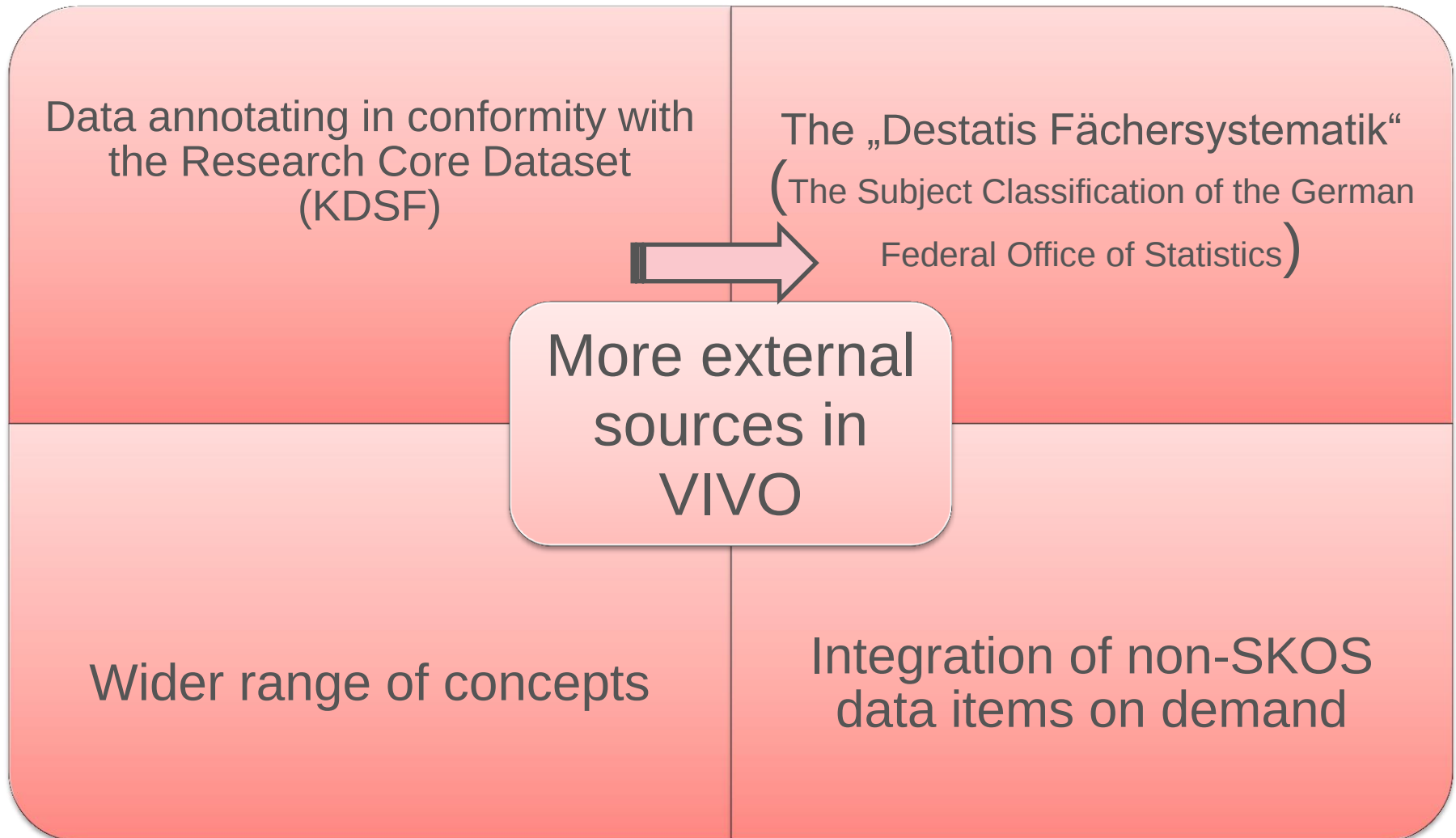
SPARQL Endpoint

Linked Data

Search Service

Label (Alternate Labels)	Definition	Best Match
☐ Linked data (Open linked data, Opendata, Linked, Linked open data, Data, Linked,...)	No definition provided.	✓

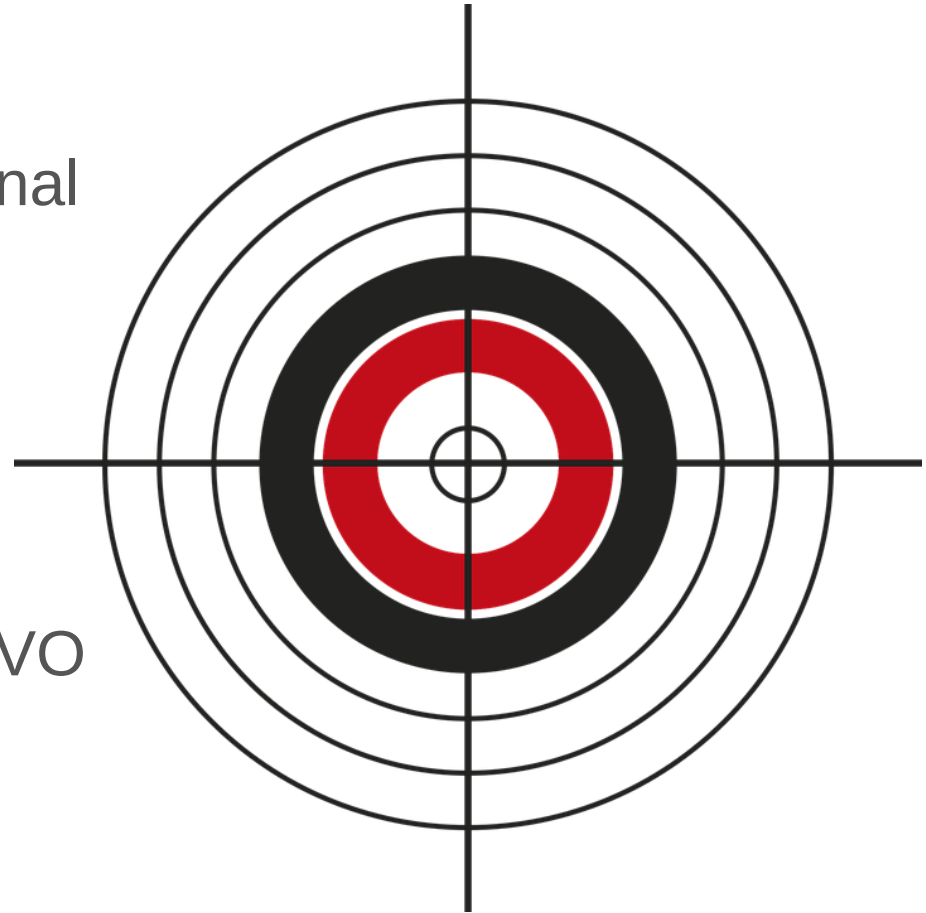
Add Selected Concept



Project Goal

Extension of the scope of external vocabularies and sources in VIVO, integrated via lookup services for:

- TIB Staff
- German users of (KDSF)VIVO
- Other interested parties



- **SKOS-based vocabularies:**

1. **Skosmos** for the sources like „Destatis Fächersystematik“ (The Subject Classification of the German Federal Office of Statistics)

- Initially in non-machine-readable form
- Transformed into **skos:ConceptScheme** by means of **Skosify**
- For an access both by a VIVO application and users on the web

2. **Direct access** to subject authorities, e. g. STW Theasaurus for Economics, already available in SKOS, having a public API

- **Other sources:**

- Data authorities in rdf, published on the web - in the pipeline

Why Skosmos?



- Web-Publishing tool for vocabularies in SKOS
- Skosmos is being developed at the National Library of Finland
- Open source with the MIT license
- Github community to participate submitting bug reports, requesting
- For machines: access via REST API
- User-friendly access to vocabularies:

- Browsing
- Alphabetical index
- Thematic index
- Visualization of concept hierarchy
- Multi-language interface

<http://skosmos.org/>

The screenshot shows the Skosmos web interface for the 'Fächersystematik' vocabulary. The interface is in German. On the left, there is a navigation pane with tabs for 'Alphabetisch' and 'Baumstruktur'. Below 'Alphabetisch', a list of concepts is shown, starting with 'A' and including 'Afrika', 'Afrikanistik', 'Agrar-, Forst- und Ernährungswissenschaften allgemein', etc. The main content area on the right is titled 'Informationen zum Vokabular'. It contains metadata for the 'Fächersystematik' vocabulary, including the version '2017-05-15', the URI 'http://www.w3.org/2004/02/skos/core#ConceptScheme', and a comment in German stating that the ontology is based on 'Bildung und Kultur: Personal an Hochschulen - Fächersystematik - 2015'. Below this, there are two tables: 'Anzahl Ressourcen nach Typ' and 'Zahl der Begriffe nach Sprache'. The first table shows 1 resource of type 'Begriffe'. The second table shows 736 concepts in German, 0 in English, and 0 in other languages. At the bottom, there is a link to 'Download des Vokabulars in SKOS/RDF: RDF/XML'.

Architecture

- PHP and JavaScript
- Open source libraries
- Plugins: Composer, EasyRDF, Twig, jQuery, jsTree, Bootstrap, typeahead.js, URI.js

Requirements:

- one or more SKOS vocabularies
- PHP capable web server
- SPARQL triple store (we recommend Apache Jena Fuseki with jena-text)

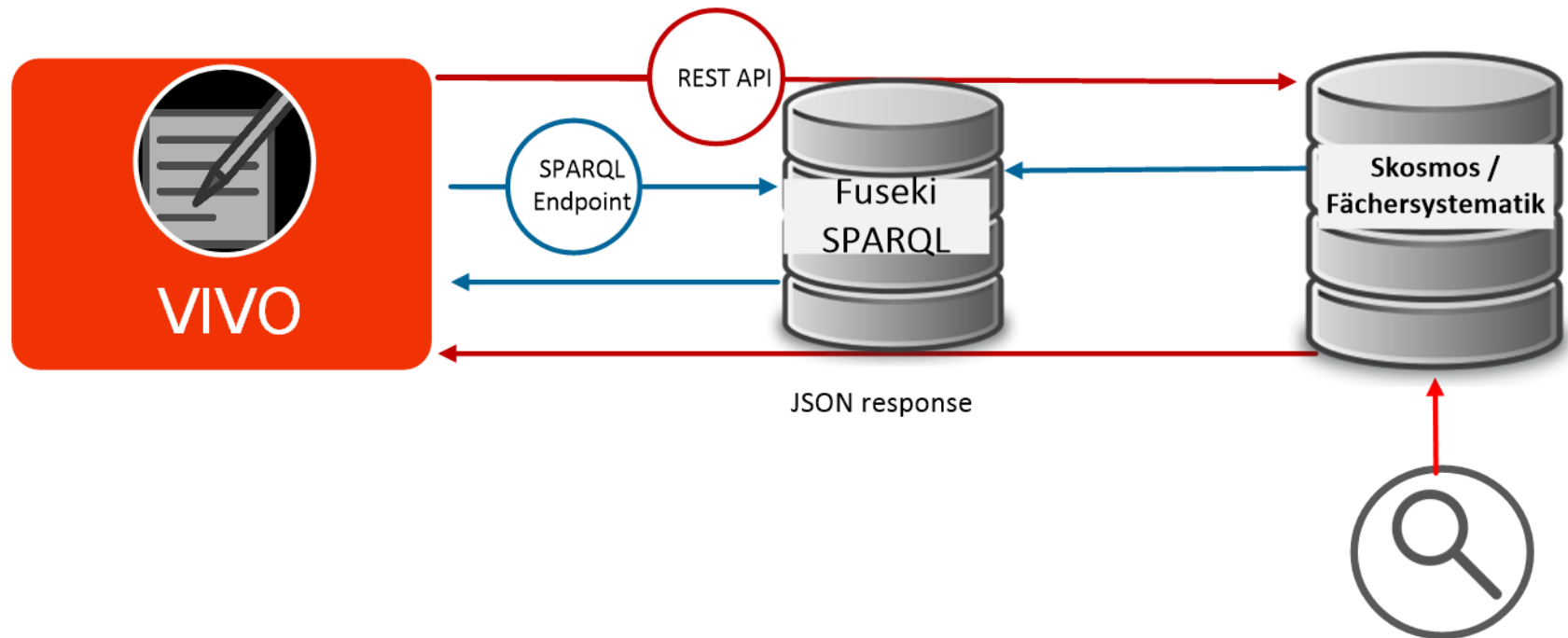
Access:

- REST-style API
- Linked Data access

Technical Implementation: VIVO <-> Skosmos



„Destatis Fächersystematik“ in VIVO
via Skosmos and REST API



- **Query URL:**

https://labs.tib.eu/skosmos/rest/v1/search?query=Bioinformatik*&vocab=faechersystematik&lang=de

- **JSON response:**

```
{"@context":{"skos":"http://www.w3.org/2004/02/skos/core#","isothes":"http://  
Vpurl.org/Viso25964/skos-  
thes#","onki":"http://schema.onki.fi/Onki#","uri":"@id","type":"@type","results":{"  
@id":"onki:results","@container":"@list"},"prefLabel":"skos:prefLabel","altLabel":  
"skos:altLabel","hiddenLabel":"skos:hiddenLabel","@language":"de"},"uri":"","res  
ults":[{"uri":"https://onto.tib.eu/destf/cs/3540","type":["skos:Concept"],"localna  
me":"3540","prefLabel":"Bioinformatik","lang":"de","notation":"3540","vocab":"fae  
chersystematik"]}]}
```

Fächersystematik in Skosmos: Demo

skosmos

Vocabularies About Feedback Help | auf Deutsch

Fächersystematik

Content language German x Search

Alphabetical

Hierarchy

A

Afrika

Afrikanistik

Agrar-, Forst- und Ernährungswissenschaften allgemein

Agrar-, Forst- und Ernährungswissenschaften allgemein

Agrar-, Forst- und Ernährungswissenschaften, Veterinärmedizin

Agrarbiologie

Agrartechnik

Agrarwissenschaften allgemein

Agrarwissenschaften, Lebensmittel- und Getränketechnologie

Akademische Selbstverwaltung

Akademisches Auslandsamt

Aktionen, Performance, Environment, Fotografie

Allg. und vergleichende Literatur- und Sprachwissenschaft allgemein

Allg. und vergleichende Literaturwissenschaft

Allg. und vergleichende Sprachwissenschaft

Allgemeine Elektrotechnik

Allgemeine Hochschulverwaltung

Allgemeine innere Verwaltung

Allgemeine Kulturwissenschaft

Allgemeine und kognitive Psychologie

Allgemeine und vergleichende Literatur- und Sprachwissenschaft

Allgemeinmedizin

Alte Geschichte

Altes Testament (evang. Th.)

Altkatholische Theologie

Altorientalistik

Altphilologie (klass. Philologie)

Altphilologie allgemein

Ambulanz, Konsiliardienst, soweit nicht fachlich zuzuordnen

Amerikanistik

Analytische Chemie

Anatomie

Anatomie, Embryologie und Histologie

Andrologie und Haustierbesamung

Angewandte Biotechnologie (Agrar-, F.- u. E.-wiss.)

Angewandte Informatik

Angewandte Kunst

Angewandte Maschinenbautechnik (Agrar-, F.- u. E.-wiss.)

Angewandte Mathematik

Angewandte Naturwissenschaften (Agrar-, F.- u. E.-wiss.)

Angewandte Sprachwissenschaft, berufsbezogene Fremdsprachenausbildung

Angewandte Verfahrenstechnik (Agrar-, F.- u. E.-wiss.)

Antiktitik

Vocabulary information

PREFERRED

Personal an Hochschulen - Fächersystematik

TERM

VERSION

2017-05-15

TYPE

<http://www.w3.org/2004/02/skos/core#ConceptScheme>

DESCRIPTION

Diese Ontologie basiert auf "Bildung und Kultur: Personal an Hochschulen - Fächersystematik - 2015; Stand: Berichtsjahr 2015; Statistisches Bundesamt (Destatis), 2015"; in SKOS konvertiert von Tatiana Walther unter Mitwirkung von Christian Hauschke und Anna Kasprzik (Technische Informationsbibliothek (TIB) Hannover). Die Bezeichnungen und Inhalte der Fächersystematik blieben unverändert.
© Statistisches Bundesamt (Destatis), 2017

URI

<https://onto.tib.eu/destf/cs/>

Resource counts by type

TypeCount

≡

Term counts by language

Language	Preferred terms	Alternate terms	Hidden terms
German	736	0	0

Download this vocabulary as SKOS/RDF: RDF/XML

https://labs.tib.eu/skosmos/faechersystematik/en?clang=de

Fächersystematik in VIVO: Demo


connect • share • discover


Index Site Admin root

Home People Organizations Research Events Capability Map

Search

Admin Panel
Edit this Individual

Resource URI: <https://onto.tib.eu/destf/cs/3540>

Photo

Resource URI

Biomformatik

Concept

Overview Identity View All

research area of

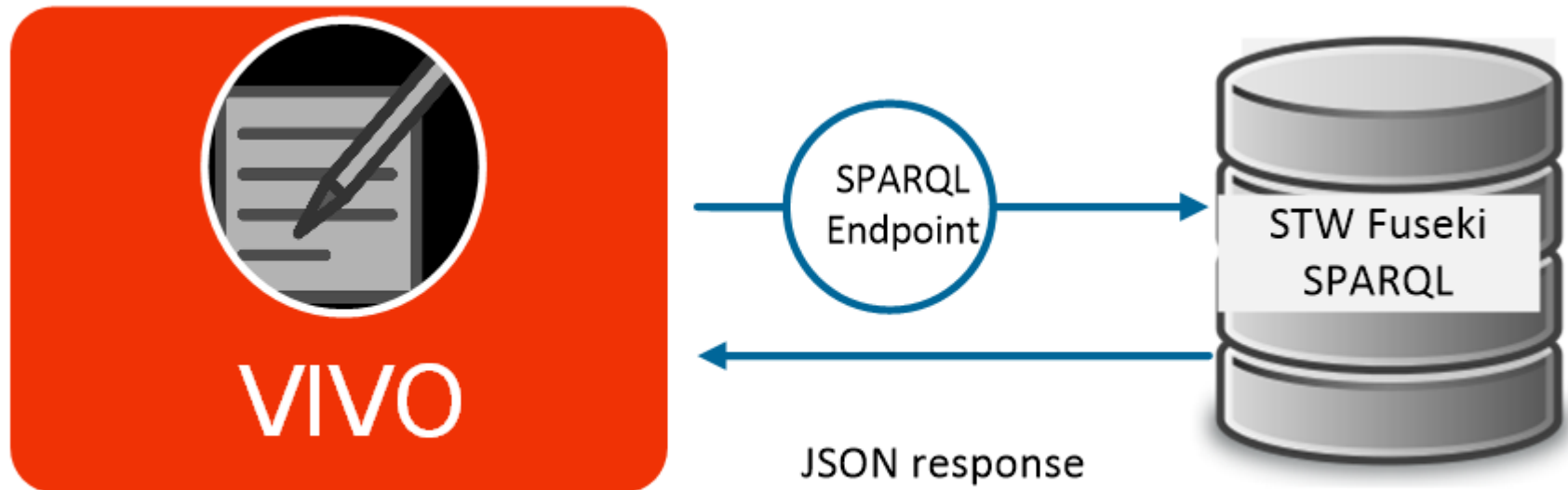
Walther, Tatiana

broader concept

narrower concept

Technical Implementation: VIVO <-> External Source

- **STW Thesaurus for Economics** in VIVO
- <http://zbw.eu/stw/version/9.04/about.en.html>
- public SPARQL Endpoint



- Query in the java class:

```
String service = "http://zbw.eu/beta/sparql/stw/query";

String sparqlQueryString2 = "PREFIX skos: <http://www.w3.org/2004/02/skos/core#>"
    + "PREFIX text: <http://jena.apache.org/text#>"
    + "SELECT ?subject ?prefLabel WHERE { VALUES ( ?searchString ?language )"
    + "{ ( \"\" + keyword + \"\" \"\" + queryLanguage + \"\" ) }"
    + "?subject text:query ?searchString ; skos:prefLabel ?prefLabel ."
    + "filter (lang(?prefLabel) = ?language)}";
```

- JSON response:

```
"results": {
  "bindings": [
    {
      "subject": { "type": "uri" , "value": "http://zbw.eu/stw/descriptor/29081-1" } ,
      "prefLabel": { "type": "literal" , "xml:lang": "en" , "value": "Data" }
```

Demo: STW Thesaurus For Economics in VIVO


connect • share • discover

[Home](#)
[People](#)
[Organizations](#)
[Research](#)
[Events](#)
[Capability Map](#)

[Search](#)

[Admin Panel](#)
[Edit this individual](#)
Verbose property display is **off** | [Turn on](#)

Resource URI: <https://vivo.tib.eu/110Beta/individual/n4242>



Walther, Tatiana

Preferred Title

Faculty Member

Positions

Research Areas

[Bibliothek](#) (Faechersystematik)
 [Bioinformatik](#)
[Daten](#) (Standard Thesaurus Wirtschaft)
 [Logik](#) (Faechersystematik)
 [Mathematik, Naturwissenschaften allgemein](#)
[Medienwissenschaften](#) (Faechersystematik)

[Overview](#)
[Affiliation](#)
[Publications](#)
[Research](#)
[Teaching](#)
[Service](#)
[Background](#)

[Contact](#)
[Identity](#)
[View All](#)

overview

geographic focus

Contact Info

Primary Email

Additional Emails

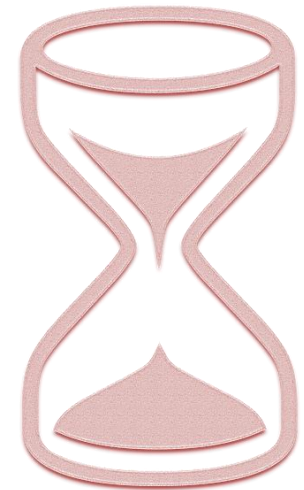
Phone

Websites

©2017 VIVO Project | Terms of Use | Powered by VIVO | Version 1.10.0-beta2
 [About](#)
[Contact Us](#)
[Support](#)

Challenges

- How to keep integrated concepts and other objects, especially e. g. organizations, up-to-date in VIVO, if they were changed in the source vocabulary?
- **Ideas:**
 1. CronJobs: Executing Cronjob nightly/ weekly/ monthly
 2. Update function, which keeps concepts/objects up-to-date and is executed every time, e.g. when Welcome page is loaded



- **Integration of data in RDF from external authorities:**
 - Entities like organizations, events, places, languages etc. in rdf
 - Possible sources: Wikidata, GND (Integrated Authority File of the German National Library), Open PIIR (Open Persistent Institutional Identifier Registry)
- **Standardized dynamic integration of external sources in VIVO via site admin and editor UI:**
 - External sources added over time by different committers
 - New integrations – challenging, demand changes in the source code

Dynamic integration of external sources in VIVO: Example Site Admin UI



Index

Site Admin

root



Home People Organizations Research Events Capability Map

Search

external sources [Add new Source](#)

Filter by source



Delete

| 4 sources |

25



external sources per page | 1 of 1

Search source

☐ Label	↕ URL from external source	↕ Link to external source	↕ Status ↕
☐ STW	http://zbw.eu/beta/sparql/stw/query	http://zbw.eu/beta/sparql-lab/about/	Online
☐ Faechersystematik	http://skosmos22.osl.tib.eu/skosmos/rest/v1/	https://labs.tib.eu/info/projekt/skosmos/	Online

Delete

| 4 sources |

25



external sources per page | 1 of 1

Dynamic integration of external sources in VIVO: Example Editor UI



VIVO | connect • share • discover



Index

Site Admin

root



Home People Organizations Research Events Capability Map

Search

Manage Concepts

Concept (Type)	Vocabulary Source
ta	Standard Thesaurus Wirtschaft Remove
ie Partnerschaft	Remove
ormationsaustausch	Remove

External Vocabulary Services

- ☒ EuroVoc external Skosmos (EuroVoc, the EU's multilingual thesaurus German National Library of Science and Technology(TIB) Labs Skosmos)
- ☐ Fachersystematik external Skosmos (German National Library of Science and Technology(TIB) Labs Skosmos)
- ☐ Fachersystematik internal Fuseki (German National Library of Science and Technology(TIB) Labs Fuseki)
- ☐ Fachersystematik internal Fuseki (German National Library of Science and Technology(TIB) internal Fuseki)
- ☐ GEMET (General Multilingual Environmental Thesaurus)
- ☐ LCSH (Library of Congress Subject Headings)
- ☐ Standard Thesaurus Wirtschaft (ZBW)
- ☐ UMLS (Unified Medical Language System)

Select a Response Format

JSON

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX odp: <http://data.europa.eu/euodp/ontologies/ec-odp#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
select distinct ?g ?o where { graph ?g { ?s dc:title ?o. filter regex(?o, 'Statistics', 'i') } } LIMIT 10
```

Select a Response Language

EN

URL executed

```
/foaf/0.1/> select distinct ?g ?o where { graph ?g { ?s dc:title ?o. filter regex(?o, 'Statistics', 'i') } } LIMIT 10&format=text/html
```

Run Query

[Return to Profile Page](#)

Any questions, comments or ideas?

Thank you for your attention!

Images

Thumb down - <https://pixabay.com/de/abneigung-hand-daumen-nach-unten-157252/>

Thumb up - <https://pixabay.com/de/hand-wie-daumen-bis-best%C3%A4tigen-157251/>

Target - <https://pixabay.com/de/zielscheibe-ziel-bogenschie%C3%9Fen-2304567/>

LD4L logo - <https://www.ld4l.org/>

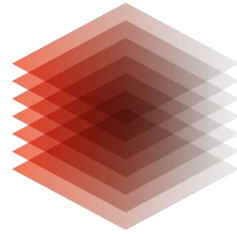
D-Space logo - <https://wiki.duraspace.org/display/DSDOC5x/Discovery>

VIVO logo - <https://wiki.duraspace.org/download/attachments/33949059/vivo-web-large-v2.gif?api=v2>

National Library of Finland - <https://www.kansalliskirjasto.fi/en>

Sand glas - <https://pixabay.com/de/sanduhr-uhr-symbol-sand-gold-2986417/>

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

Contact:

Tatiana Walther
tatiana.walther@tib.eu

Martin Barber
martin.barber@tib.eu