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
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

- **Feasable ways of integration**: ingesting data dumps vs. direct access to external sources via lookup services
VIVO in Brief

- 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](REST_API.png)

![SPARQL Endpoint](SPARQL_Endpoint.png)

Linked Data

<table>
<thead>
<tr>
<th>Label (Alternate Labels)</th>
<th>Definition</th>
<th>Best Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked data</td>
<td>No definition provided.</td>
<td>✔️</td>
</tr>
</tbody>
</table>

(Open linked data, Opendata, Linked, Linked open data, Data, Linked, ...)
Motivation

Data annotating in conformity with the Research Core Dataset (KDSF)

The „Destatis Fächersystematik“ (The Subject Classification of the German Federal Office of Statistics)

More external sources in VIVO

Wider range of concepts

Integration of non-SKOS data items on demand
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
Approaches

• **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/
Technical Implementation: Skosmos

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
Technical Implementation: Fächersystematik in VIVO

- **Query URL:**

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

- **JSON response:**

  ```json
  {"@context":{"skos":"http:\/\/www.w3.org\/2004\/02\/skos\/core#","isothes":"http:\/\/
purl.org\/iso25964\/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":"","results":[{"uri":"https:\/\/onto.tib.eu\/destf\/cs\/3540","type":["skos:Concept"],"localname":"3540","prefLabel":"Bioinformatik","lang":"de","notation":"3540","vocab":"fächer
csystematik"}]
  ```
Fächersystematik in Skosmos: Demo
Fächersystematik in VIVO: Demo
Technical Implementation: VIVO <-> External Source

- STW Thesaurus for Economics in VIVO
- public SPARQL Endpoint
Technical Implementation STW in VIVO

- Query in the java class:

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

```

- JSON response:

```json
"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
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
Future Plans

• 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
Dynamic integration of external sources in VIVO: Example Editor UI
Any questions, comments or ideas?

Thank you for your attention!
Images


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


Contact:

Tatiana Walther
tatiana.walther@tib.eu

Martin Barber
martin.barber@tib.eu