

BBT Browser & Federated thesauri viewer aka ACDH Vocab

Ksenia Zaytseva, Matej Ďurčo @ACDH-OEAW
BBT Workshop, 28.9. 2018, Heraklion



Outline

- BBT BRowser / Vocabulary repository - the tool
- Visualization component
Requirements, prototype, plans
- Data quality and curation

BBT browser aka ACDH Vocab

- <https://vocabs.dariah.eu/en/>
- Instance of [SKOSMOS](#) ([documentation](#))
- Php + triplestore (Apache Jena Fuseki by default)
- Main Features:
 - follows SKOS data model; Dublin core; some of rdfs, owl , foaf properties
 - supports multilingual thesauri
 - provides RDF/XML, Turtle, JSON-LD serialization for each concept; download for a whole vocabulary
 - statistics (how many concepts and their labels in each available language)
 - SKOS vocabularies should be well-formed to use SKOSMOS functionality fully (important to have reverse relationships - use [skosify](#) to pre-process/fix SKOS)
 - REST API
 - SPARQL endpoint <https://vocabs-sparql.acdh-dev.oeaw.ac.at/>

Components



Browser

ACDH Vocabularies

Vocabularies About Feedback Help

A Vocabulary Repository Service by ACDH

In many areas of scholarly work, controlled vocabularies (gazetteers) serve as a stable reference for resources and for ensuring interoperability of their data. The ACDH provides a vocabulary repository service for the maintenance and publication of vocabularies and taxonomies.

[Vocabularies](#)

SPARQL endpoint

ACDH Vocabularies SPARQL

About

ACDH Vocabularies

from all English

Skosmos Vocabulary Categories

GENERAL CONCEPTS

[Backbone Thesaurus](#)

CULTURAL HISTORY

[travel!digital Thesaurus](#)

ARCHAEOLOGY

[DEFC Thesaurus](#)

MUSIC

[musical instrument classification](#)

ARCHE

[ARCHE access restrictions](#)

[ARCHE life cycle status](#)

[ARCHE Resource Type Category](#)

Example queries

Fetch all triples

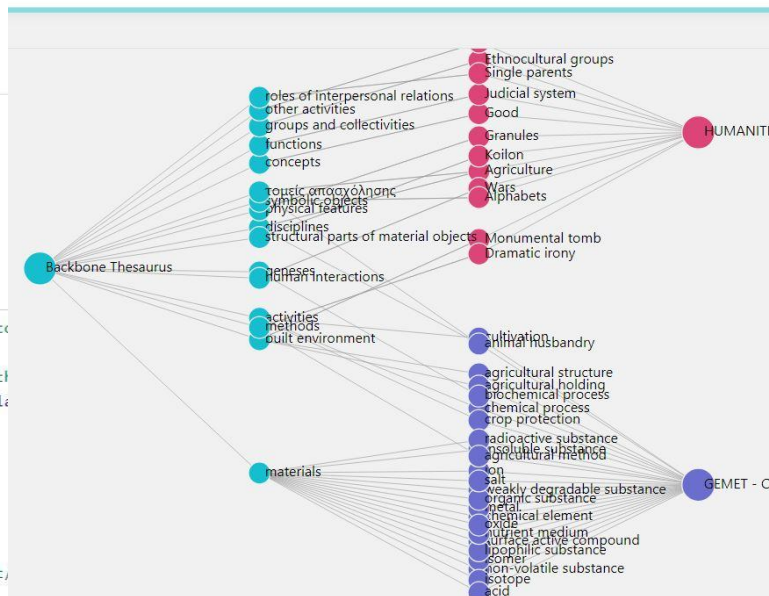
Fetch all skos:broadMatch

Query

```
1 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
2 PREFIX dc: <http://purl.org/dc/elements/1.1/>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 SELECT ?c ?c_label ?r ?c2_label ?c2 ?s2 ?s2_label
5 WHERE
6 {
7   ?c a skos:Concept;
8     skos:prefLabel ?c_label.
9   ?c2 a skos:Concept;
10     skos:prefLabel ?c2_label.
11   ?c skos:inScheme <https://vocabs.dariah.eu/bbt/ConceptScheme/Backbone_Thesaurus>
12   ?c2 skos:inScheme ?s2.
13 }
14 OPTIONAL { ?s2 dc:title | rdfs:label ?s2_label. }
15 FILTER ( <https://vocabs.dariah.eu/bbt/ConceptScheme/Backbone_Thesaurus> != ?s2 )
```

ACDH Vocabularies visualization

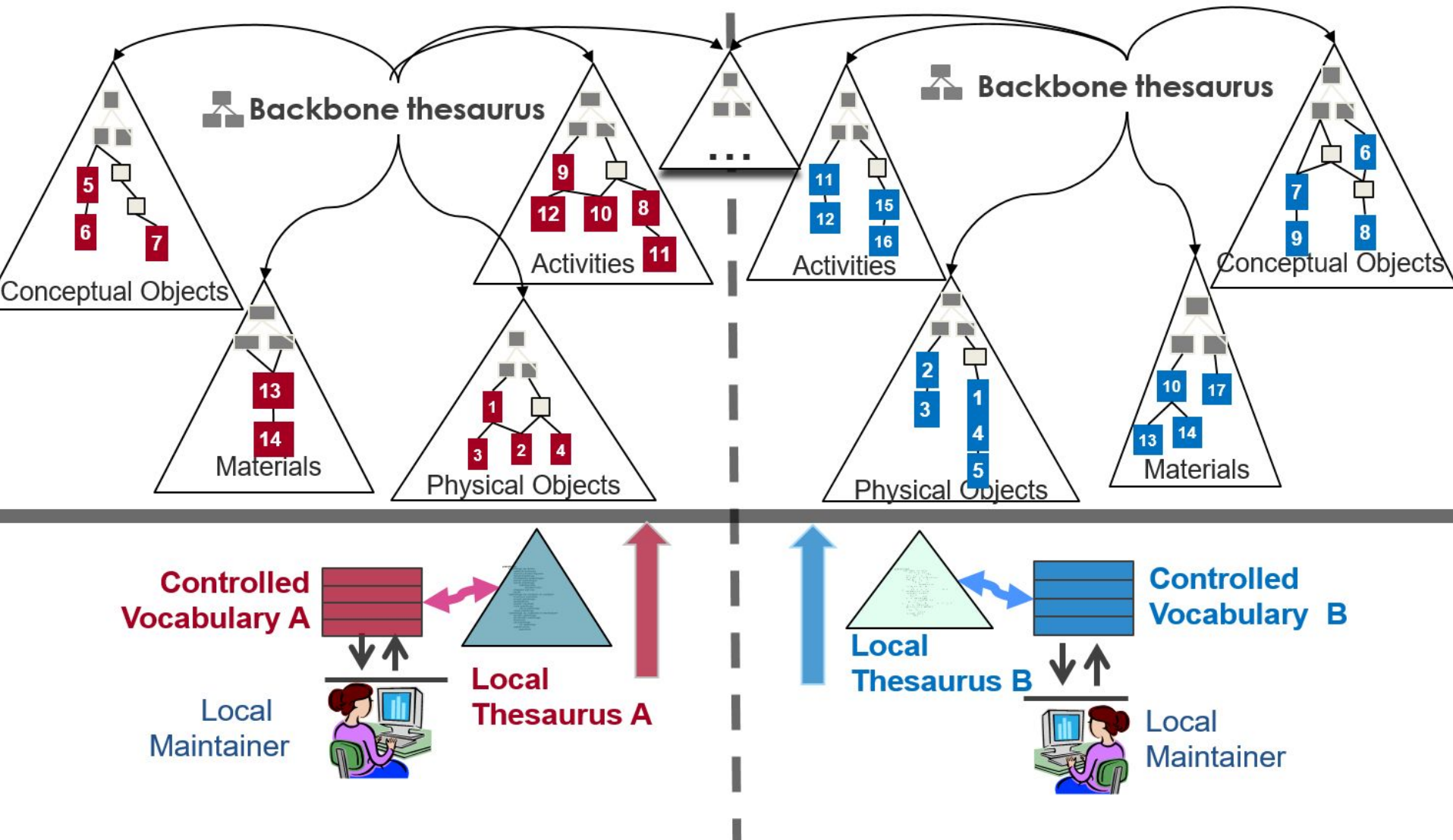
Visualisation



Visualisation



Motivation/Inspiration:



→ Requirements:

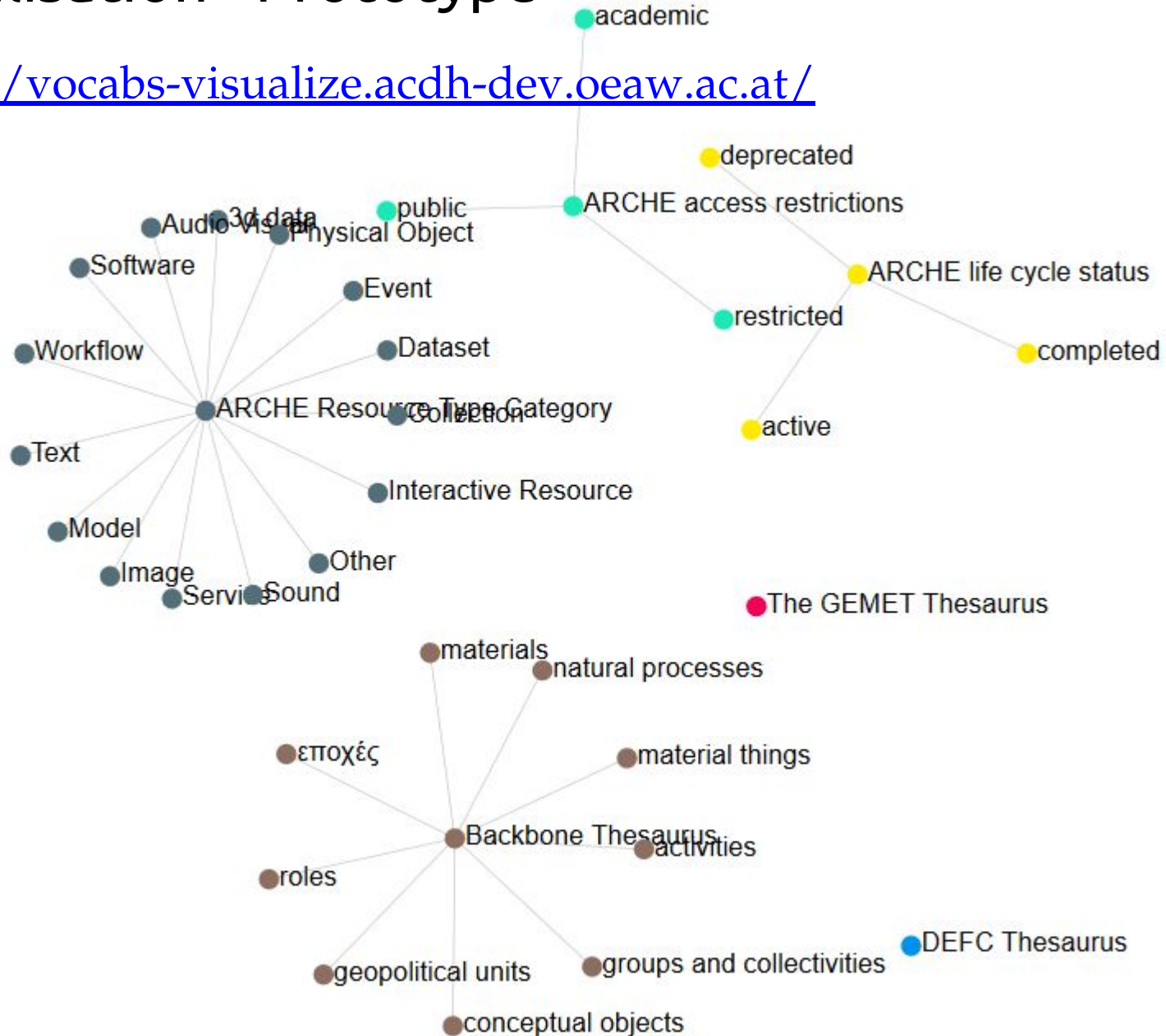
- ◆ Allow to select and show multiple concept schemes as a concept tree/hierarchy
- ◆ Show the relations between concepts from different concept schemes
- ◆ Show the “neighbourhood” of a concept,
i.e. parents, children, related concepts from other concept schemes
(as separate widget in the concept-detail view)

→ Practical issues:

- ◆ Concept schemes can be too large to sensibly display in one go
=> need to restrict the displayed sub-graph/tree
- ◆ Relations may be between concepts deeper in the hierarchy
=> dedicated query + view that picks the related concepts
- ◆ SKOSMOS stores every vocabulary as separate graph in the triple store, SPARQL queries between the graphs not possible
=> create a separate graph “vocabs_all”, accommodating all vocabularies
- ◆ When not all concepts have English labels on which language to display them?
=> filter only English labels now

Visualisation - Prototype

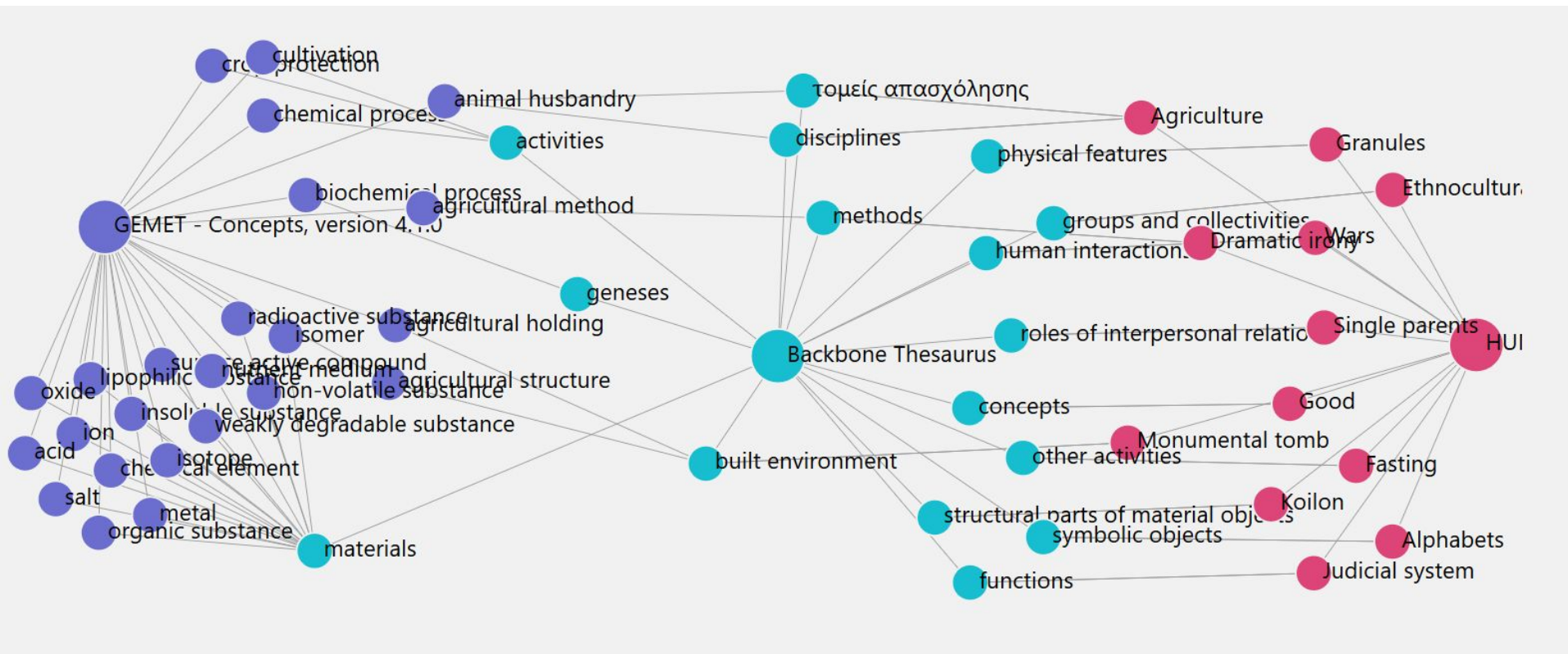
<https://vocabs-visualize.acdh-dev.oeaw.ac.at/>



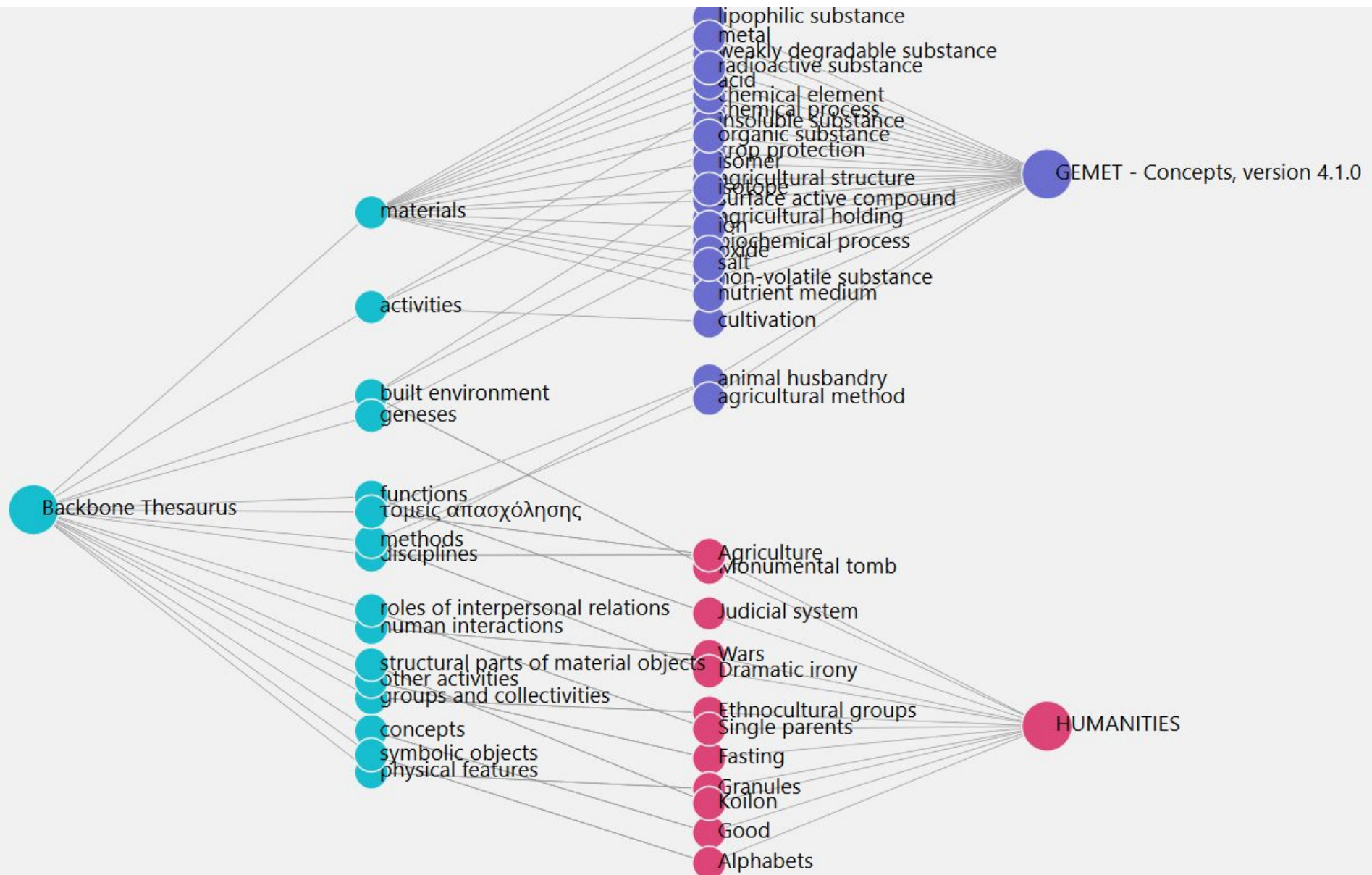
Visualisation - Prototype



Subselection of connected concepts



Visualisation - Prototype - directed view





Planned features

- Select concept context
And show their neighbourhood
- Include graph widget in the concept detail view
- Link node to concept detail view in browser
- Expand node to show children
- Over page with statistics
How many graphs, how many connected concepts
- Language switch
- Typed edges (indicate skos relation)

SPARQL queries behind visualisation



```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?c ?c_label ?r ?c2_label ?c2 ?s2 ?s2_label
WHERE
{
  ?c a skos:Concept;
    skos:prefLabel ?c_label.
  ?c2 a skos:Concept;
    skos:prefLabel ?c2_label.
  ?c ?r ?c2.
  ?c skos:inScheme <https://vocabs.dariah.eu/bbt/ConceptScheme/Backbone_Thesaurus>.
  ?c2 skos:inScheme ?s2.

  OPTIONAL {?s2 dc:title | rdfs:label ?s2_label.}
  FILTER (<https://vocabs.dariah.eu/bbt/ConceptScheme/Backbone_Thesaurus> != ?s2)
  FILTER (LANG(?c_label)="en")
  FILTER (LANG(?c2_label)="en")
}
ORDER BY ASC(str(?c_label))
```



Updating mechanism

- BBTalk will automatically generate new versions of BBT & aux files
- BBTbrowser will fetch them and update its local triple store
- Structured in separate files:
 - http://backbonethesaurus.eu/BBTalk/bbtFiles/BBT_official.rdf
Main BBT file containing all BBT concepts
Updated e.g. on releasing a new BBT version
 - http://backbonethesaurus.eu/BBTalk/bbtFiles/BBT_connected.rdf
All connections to local thesauri
Updated e.g. on releasing a new BBT version and on creating a new connection, or periodically, and also on admin-request
 - Connected local thesauri
http://backbonethesaurus.eu/BBTalk/bbtFiles/BBT_connected_thesauri.xml

Issues with data quality

General consistency issues (SKOS proper):

- Concept URLs (should be resolvable, “cool URIs”, no IP-addresses)
- Inconsistent Language tags
- More than one prefLabel in one language
- Duplicated relations between concepts
(both skos:narrower and skos:narrowerMatch)

=> Introduce automated validity checks before ingesting



Issues with data quality

Common features required for proper visualization:

- Concept Scheme: URI, `rdf:type skos:ConceptScheme` and label
- `skos:hasTopConcept` and `skos:topConceptOf`
- Concept: one `skos: prefLabel` in a given language
- All Concept have `skos:inScheme` property pointing to main Concept Scheme URI