

SW4CH 2018

# VisTA: Visual Terminology Alignment Tool for Factual Knowledge Aggregation

A. Axaridou, K. Konsolaki, M. Theodoridou, A. Kozlov, P. Haase & M. Doerr

## Contents of presentation

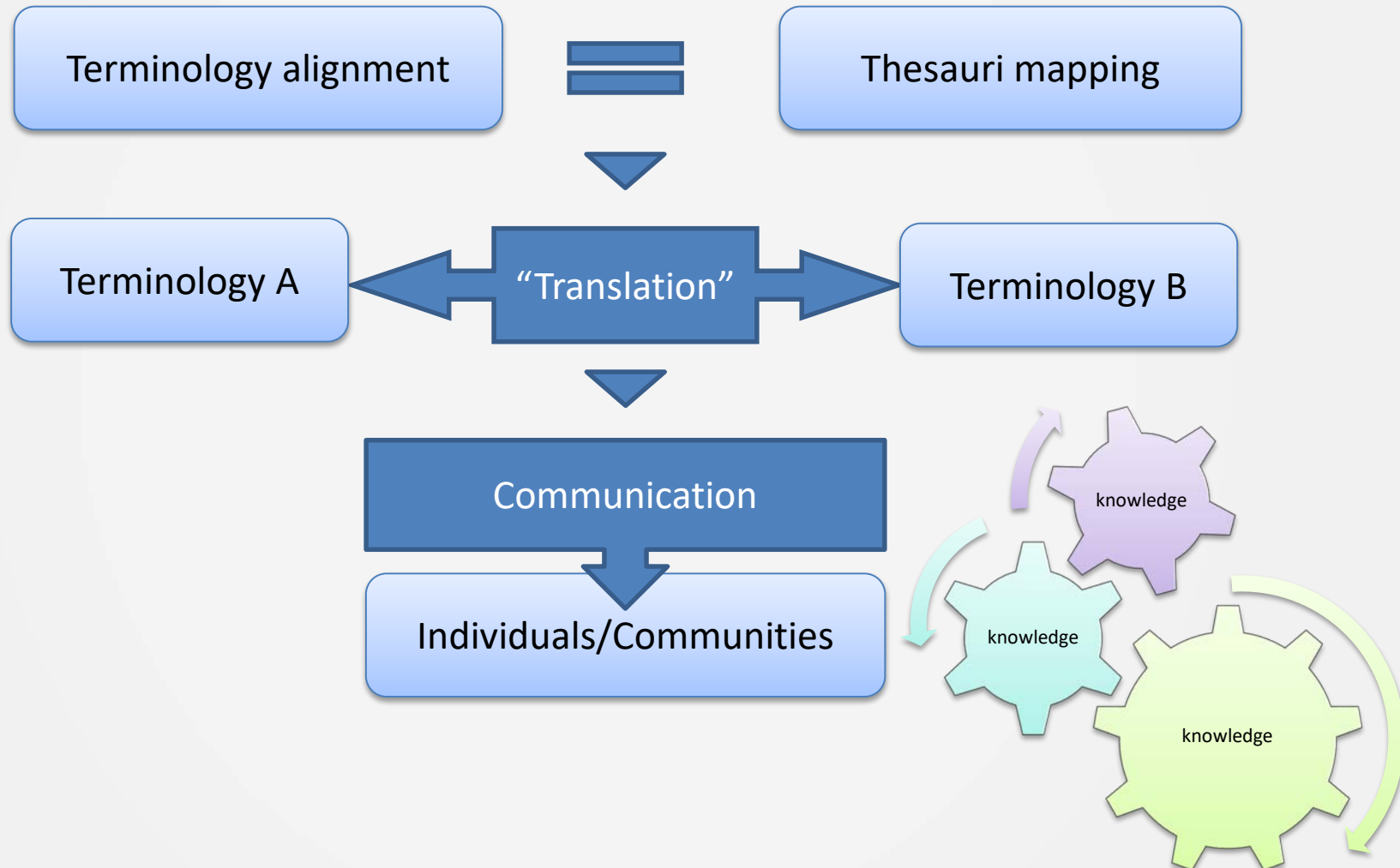
- Overview
- A short discussion about Alignment
- Motivation
- Related work
- The Alignment problem in VisTA
- Key features
- Alignment rules
- Alignment result
- VisTA GUI & features
- Configuration
- Conclusions

This work is the result of collaboration between the British Museum, metaphacts and FORTH in the context of the ResearchSpace project.

## Overview

- **Exact** (not approximate) alignment for RDF/SKOS terminologies => enables factual knowledge aggregation
- A simple and friendly **web-based** user interface for the alignment between two terminologies
- **Visualizes** the terminology hierarchies, enables the **interactive** alignment process, and presents the alignment result
- Component of the **Synergy Reference Model** (an initiative of the CIDOC CRM Special Interest Group for manipulation of data provisioning and aggregation processes)

# What is Terminology Alignment



## About Alignment 1/2

- The broader area of Alignment comprises: **schema matching, schema mapping, ontology alignment**
- Alignment is: The process of creating **correspondences** (direct associations) between terms/concepts of different vocabularies. E.g. “**alignment**” may also be known as “**matching**”
- Proposed methods: applying **automatic** or **semi-automatic** procedures based on configurable algorithms and workflows. Many approaches usually deal with **literal matching** of the terms along with a **structural matching**
- Focusing on: **large** datasets/schemas/ontologies/vocabularies
- Challenge: the development of approaches trying to achieve results of as high as possible **precision** and **recall**

## About Alignment 2/2

Takes place (usually)

- between two terminologies

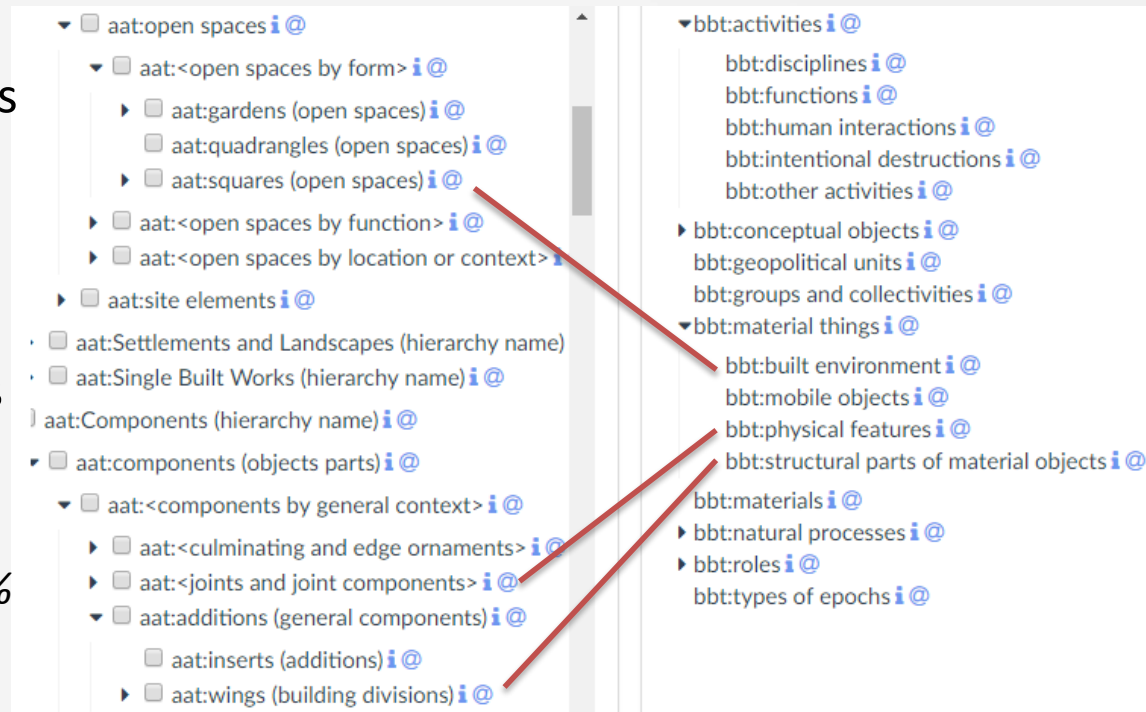
Results in:

- the **estimation of the similarity** among the terms

*e.g.*

*Sim (member, remember) = 87%*

*Sim (member, membership) = 75%*



Even if more accurate results...

The question is:

- How a **percentage value** can **factually** contribute to a **semantic knowledge aggregator**?  
(What does a percentage value semantically mean?)

## Motivation

- The approximate results of estimated similarity among terms, give no semantics about the similar terms
- To make the Alignment result useful to a knowledge aggregator **specific semantic relations** have to be defined at least between the most “similar” terms
- Last phase of an approximate alignment: the **human convention** (manual process). The only method preserving precision with high recall.
- **Lack** of GUI for supporting visual interactive alignment. Existing GUIs operate usually in 2 separated steps:
  - a **configuration phase** followed by a **rendering phase** of the result

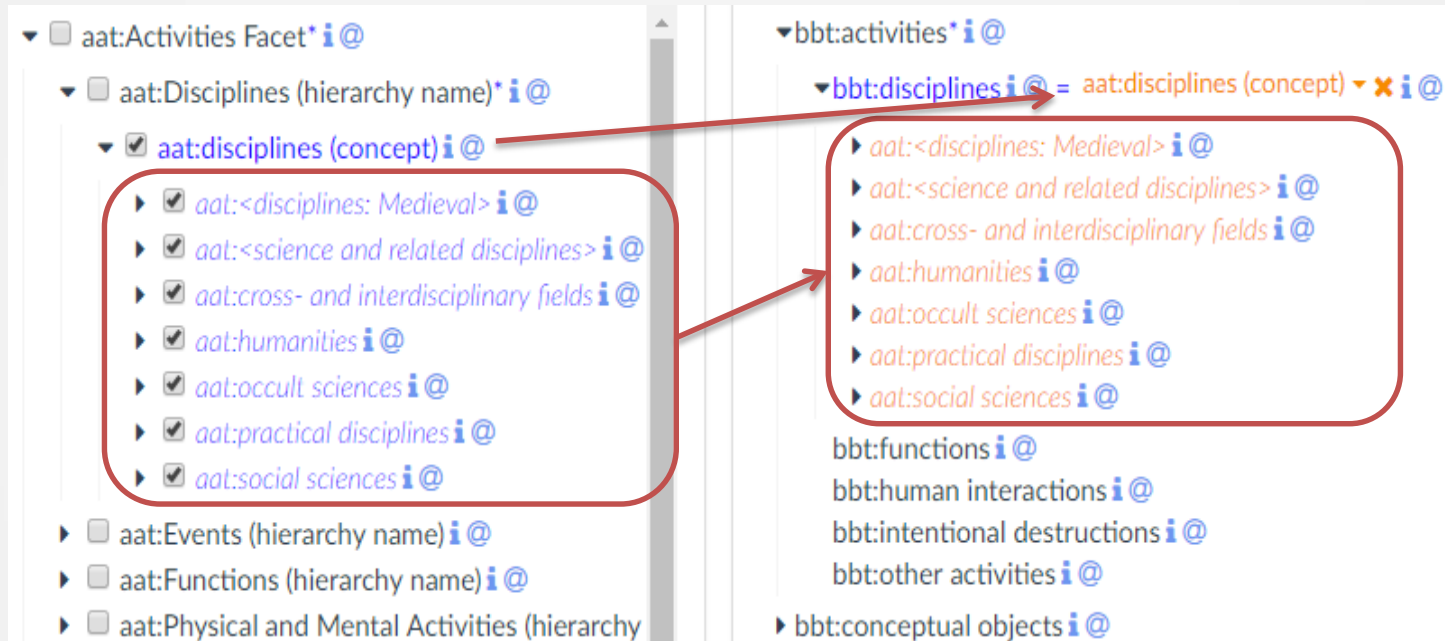
## Related work

Method	Supportive GUIs	Input size	Output accuracy	Exploitability of result
Automatic	Yes, for configs	Large	Approximate, e.g. estimation of similarity between terms	?
Semi-Automatic	Yes, for configs  Sometimes yes, for the manual phase <b>but</b> without attribution of the semantics			
Manual	VisTA. Other?	Medium/Small	Exact, i.e. define the exact relation between terms	Yes



## The alignment problem in VisTA 1/4

- Alignment of two terminologies as an *asymmetric* process, aiming to the *subordination* of a source to a target terminology.



- The process results in *n×m correspondences* among the terms of the two terminologies

## The alignment problem in VisTA 2/4

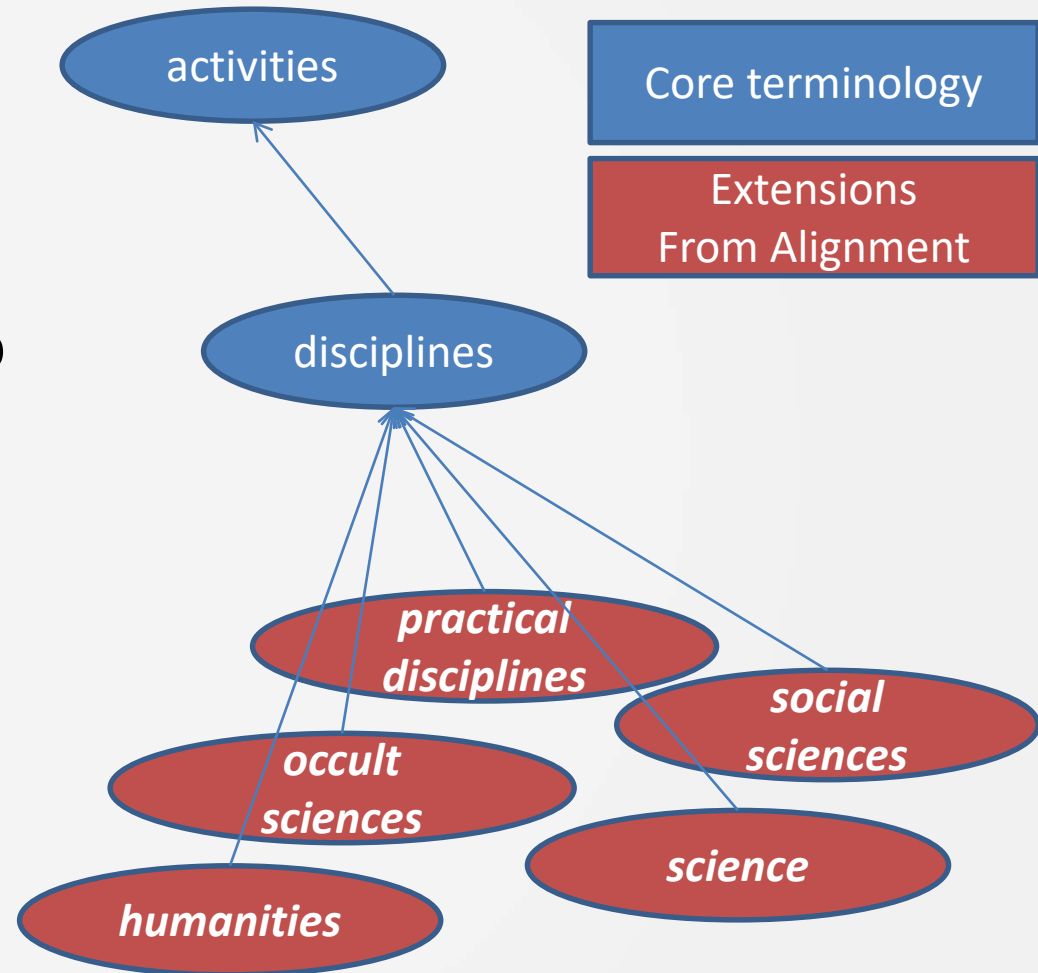
- We consider three *principles*:
  - Terminologies and the alignment result are *acyclic* graphs preserving the taxonomy *subsumption* (*is-a*)
  - The *broader* and *narrow* relations are *symmetrically inverted*
  - *Target-driven* reconciliation of the source terminology: the source conforms to the target but not the reverse

## The alignment problem in VisTA 3/4

- Goal:
  - **Integration** of different terminologies **under one target terminology used as a core terminology in a knowledge aggregator**. The core terminology gets extended with new specialized terms
  - **Empowering searching capabilities** in a semantic network, as the users of different terminologies are enabled to make queries using the **common target vocabulary together with their own familiar vocabularies**, to find more resources in their results.

## The alignment problem in VisTA 4/4

- Example of factual aggregation
  - Searching for resources related to term *disciplines*, we get all the instances related to its specialization



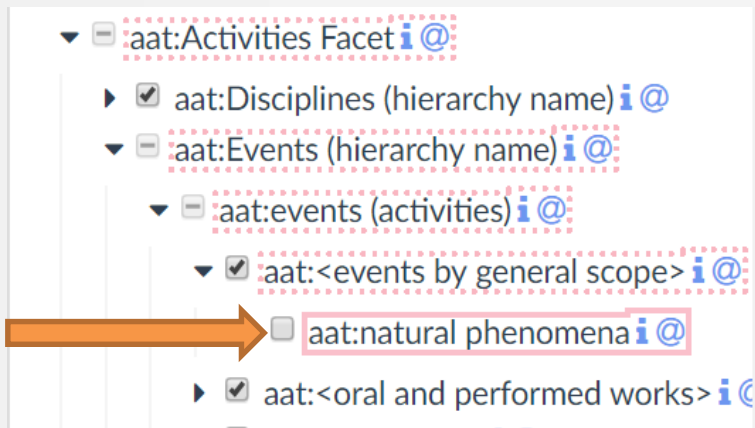
## VisTA key features

- The adaptation of the structure of the source terminology is supported by the tool: **removal of the in compliant subterms** that break the subsumption of the target (*see an example next*)
- During the process multiple-inheritance of terms, i.e. a term may have multiple parents, and subhierarchy overlaps may occur. These situations are allowed unless they break subsumption
- The extension of the target terminology is based on **broader match** and the **exact match** relations
- Alignment rules (*...more*)
- Native RDF/SKOS Alignment result (*...more*)

## Example of incompliance of the source terminology structure

### Source

#### Art & Architecture Thesaurus (AAT)



*aat: Activities Facet*

encompasses areas of endeavor,  
physical and mental actions, etc.

### Target

#### Backbone Thesaurus (BBT)

- ▶ bbt:activities
- ▶ bbt:conceptual objects
- bbt:geopolitical units
- bbt:groups and collectivities
- ▶ bbt:material things
- bbt:materials
- ▶ bbt:natural processes
- ▶ bbt:roles
- bbt:types of epochs

*bbt: activities*

classifies intentional actions

*aat: natural phenomena*



*bbt: activities*

## Alignment rules

- *Check for explicit (direct) alignment relation.* The source term must not be already aligned to the same target term, otherwise the user is informed and the process is **canceled**.
- *Check for the existence of the source term in the target tree.* When the source term to be aligned is already an *original* term of the target terminology then alignment of that term causes the change of the target terminology which is not allowed. In this case the user is informed and the process is **canceled**.
- *Check for implicit (indirect) alignment relation.* The source term may already be indirectly aligned to the target term. In this case the user gets a warning whereas the alignment is **allowed**.
- *Check for aligned descendants of the source term.* The source term may contain already aligned sub-terms. In this case the user gets a warning whereas the alignment is **allowed**.

## Alignment result

For each pair of aligned terminologies

- Alignment graph
- Native RDF/SKOS
- Contains:
  - the correspondences
  - the children hierarchy of the terms coming from the source terminology
- Searching space:
  - Alignment graph + Target terminology graph



# VisTA GUI: Import/Export page

metaphactory

sparql Account

Visual Terminology Alignment

Edit Page

Import/Export Edit Loaded alignment graph: <http://forth.gr/alignment/aat\_bbt\_2018-05-09T11:40:49.520Z>

Import terminology

Advanced Options

Please drag&drop your RDF file(s) here.

Terminologies

The repository contains 9 distinct Terminology graph(s).

Filter Results

Terminology Graph	Triples	Export	Delete
http://vocab.getty.edu/dataset/aat	>=1000	⬇	🗑
http://vocab.dariah.eu/bbt	352	⬇	🗑
http://BM/terminology	>=1000	⬇	🗑
file://nomismaScheme.rdf-19-09-2017-04-28-14	>=1000	⬇	🗑
file://Moderno_MaterialeScheme.rdf-19-09-2017-04-29-16	>=1000	⬇	🗑
http://forth.gr/nomismaMaterial	619	⬇	🗑

Alignments

The repository contains 4 distinct Alignment graph(s).

Filter Results

Alignment Graph	Triples	Export	Delete	Locked by
http://forth.gr/alignment/aat_bbt_1511176905335	2	⬇		admin
http://forth.gr/alignment/aat_bbt_1511177171270	2	⬇		admin
http://forth.gr/alignment/aat_bbt_2018-03-29T10:30:33.516Z	>=1000	⬇		admin
http://forth.gr/alignment/aat_bbt_2018-05-09T11:40:49.520Z	1	⬇	🗑	

# VisTA GUI: Edit page

The screenshot displays the VisTA GUI's 'Edit' page for a terminology alignment. The browser address bar shows 'localhost:10214/resource/forth:term-alignment'. The page header includes 'metaphactory' and 'Visual Terminology Alignment'. The main content area is titled 'Loaded alignment graph: <http://forth.gr/alignment/aat\_bbt\_2018-05-09T11:40:49.520Z>'. Below this, the 'Source Terminology' is set to '<http://vocab.getty.edu/dataset/aat>' and the 'Target Terminology' is set to '<http://vocab.dariah.eu/bbt>'. The interface is split into two panels. The left panel, 'Full Hierarchy of aat', shows a tree structure of activities and related concepts. The right panel, 'Full Hierarchy of bbt', shows a tree structure of broader transitive activities and related concepts. Both panels include a search bar and a list of terms with expandable sub-panels. A footer note at the bottom of each panel reads: 'Click on the info button next to each term to get more information about the term'.

metaphactory

sparql Account

Visual Terminology Alignment

Import/Export Edit

Loaded alignment graph: <http://forth.gr/alignment/aat\_bbt\_2018-05-09T11:40:49.520Z>

Source Terminology <http://vocab.getty.edu/dataset/aat>

Target Terminology <http://vocab.dariah.eu/bbt>

Full Hierarchy of aat gvp:broaderGeneric

- aat:Activities Facet
  - aat:Disciplines (hierarchy name)
  - aat:Events (hierarchy name)
  - aat:Functions (hierarchy name)
  - aat:Physical and Mental Activities (hierarchy name)
  - aat:Processes and Techniques (hierarchy name)
  - aat:activities (general context)
- aat:Agents Facet
- aat:Associated Concepts Facet
- aat:Brand Names Facet
- aat:Materials Facet
- aat:Objects Facet
- aat:Physical Attributes Facet
- aat:Styles and Periods Facet

Click on the info button next to each term to get more information about the term

Full Hierarchy of bbt skos:broaderTransitive

- bbt:activities
  - bbt:disciplines
  - bbt:functions
  - bbt:human interactions
  - bbt:intentional destructions
  - bbt:other activities
- bbt:conceptual objects
- bbt:geopolitical units
- bbt:groups and collectivities
- bbt:material things
- bbt:materials
- bbt:natural processes
- bbt:roles
- bbt:types of epochs

Click on the info button next to each term to get more information about the term

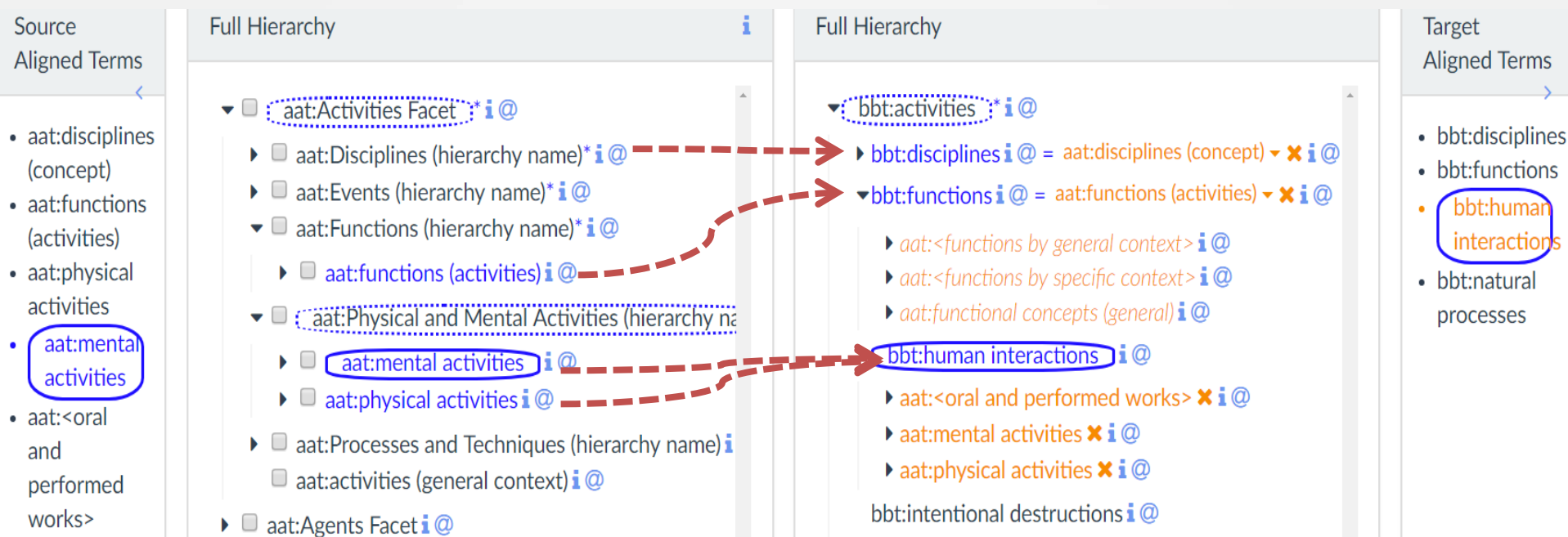
Screenshot Added

# VisTA Features 1/6

Source Aligned Terms	Full Hierarchy	Full Hierarchy	Target Aligned Terms
<ul style="list-style-type: none"> <li>• aat:disciplines (concept)</li> <li>• aat:functions (activities)</li> <li>• aat:physical activities</li> <li>• <b>aat:mental activities</b></li> <li>• aat:&lt;oral and performed works&gt;</li> </ul>	<div> <div> <input type="checkbox"/> aat:Activities Facet *i@           <ul style="list-style-type: none"> <li>▶ <input type="checkbox"/> aat:Disciplines (hierarchy name)*i@</li> <li>▶ <input type="checkbox"/> aat:Events (hierarchy name)*i@</li> <li>▼ <input type="checkbox"/> aat:Functions (hierarchy name)*i@               <ul style="list-style-type: none"> <li>▶ <input type="checkbox"/> aat:functions (activities)i@</li> </ul> </li> <li>▼ <input type="checkbox"/> aat:Physical and Mental Activities (hierarchy name)*i@               <ul style="list-style-type: none"> <li>▶ <input type="checkbox"/> <b>aat:mental activities</b> i@</li> <li>▶ <input type="checkbox"/> aat:physical activities i@</li> </ul> </li> <li>▶ <input type="checkbox"/> aat:Processes and Techniques (hierarchy name)i@               <ul style="list-style-type: none"> <li><input type="checkbox"/> aat:activities (general context)i@</li> </ul> </li> <li>▶ <input type="checkbox"/> aat:Agents Facet i@</li> </ul> </div> </div>	<div> <div> <input type="checkbox"/> bbt:activities *i@           <ul style="list-style-type: none"> <li>▶ bbt:disciplines i@ = aat:disciplines (concept) ✕ i@</li> <li>▼ bbt:functions i@ = aat:functions (activities) ✕ i@               <ul style="list-style-type: none"> <li>▶ aat:&lt;functions by general context&gt; i@</li> <li>▶ aat:&lt;functions by specific context&gt; i@</li> <li>▶ aat:functional concepts (general) i@</li> </ul> </li> <li>▼ <b>bbt:human interactions</b> i@               <ul style="list-style-type: none"> <li>▶ aat:&lt;oral and performed works&gt; ✕ i@</li> <li>▶ aat:mental activities ✕ i@</li> <li>▶ aat:physical activities ✕ i@</li> </ul> </li> </ul> </div> </div>	<ul style="list-style-type: none"> <li>• bbt:disciplines</li> <li>• bbt:functions</li> <li>• <b>bbt:human interactions</b></li> <li>• bbt:natural processes</li> </ul>

- **Visualization** of the Terminologies
  - Tree structures as **indented lists**
  - **Expandable/collapsible** nodes

## VisTA Features 2/6



- **Interactive alignment**
  - Drag'n'drop
  - Run time visualization of the alignment result on the target terminology tree

## VisTA Features 3/6

Source Aligned Terms	Full Hierarchy	Full Hierarchy	Target Aligned Terms
<ul style="list-style-type: none"> <li>• aat:disciplines (concept)</li> <li>• aat:functions (activities)</li> <li>• aat:physical activities</li> <li>• <b>aat:mental activities</b></li> <li>• aat:&lt;oral and performed works&gt;</li> </ul>	<ul style="list-style-type: none"> <li>✗ <b>aat:Activities Facet</b> *i@</li> <li>✗ ▶ <input type="checkbox"/> aat:Disciplines (hierarchy name)*i@</li> <li>✗ ▶ <input type="checkbox"/> aat:Events (hierarchy name)*i@</li> <li>✗ ▶ <input type="checkbox"/> aat:Functions (hierarchy name)*i@</li> <li>➔ aat:functions (activities)i@</li> <li>✗ ▶ <b>aat:Physical and Mental Activities (hierarchy name)</b></li> <li>➔ <b>aat:mental activities</b> i@</li> <li>➔ aat:physical activities i@</li> <li>✗ ▶ <input type="checkbox"/> aat:Processes and Techniques (hierarchy name)i@</li> <li>✗ ▶ <input type="checkbox"/> aat:activities (general context)i@</li> <li>✗ ▶ <input type="checkbox"/> aat:Agents Facet i@</li> </ul>	<ul style="list-style-type: none"> <li>✗ <b>bbt:activities</b> *i@</li> <li>➔ bbt:disciplines i@ = aat:disciplines (concept) ✗ i@</li> <li>➔ bbt:functions i@ = aat:functions (activities) ✗ i@</li> <li>➔ aat:&lt;functions by general context&gt; i@</li> <li>➔ aat:&lt;functions by specific context&gt; i@</li> <li>➔ aat:functional concepts (general) i@</li> <li>➔ <b>bbt:human interactions</b> i@</li> <li>➔ aat:&lt;oral and performed works&gt; ✗ i@</li> <li>➔ aat:mental activities ✗ i@</li> <li>➔ aat:physical activities ✗ i@</li> <li>✗ bbt:intentional destructions i@</li> </ul>	<ul style="list-style-type: none"> <li>• bbt:disciplines</li> <li>• bbt:functions</li> <li>• <b>bbt:human interactions</b></li> <li>• bbt:natural processes</li> </ul>

- **Visualization** of a Term

- States of a Term:

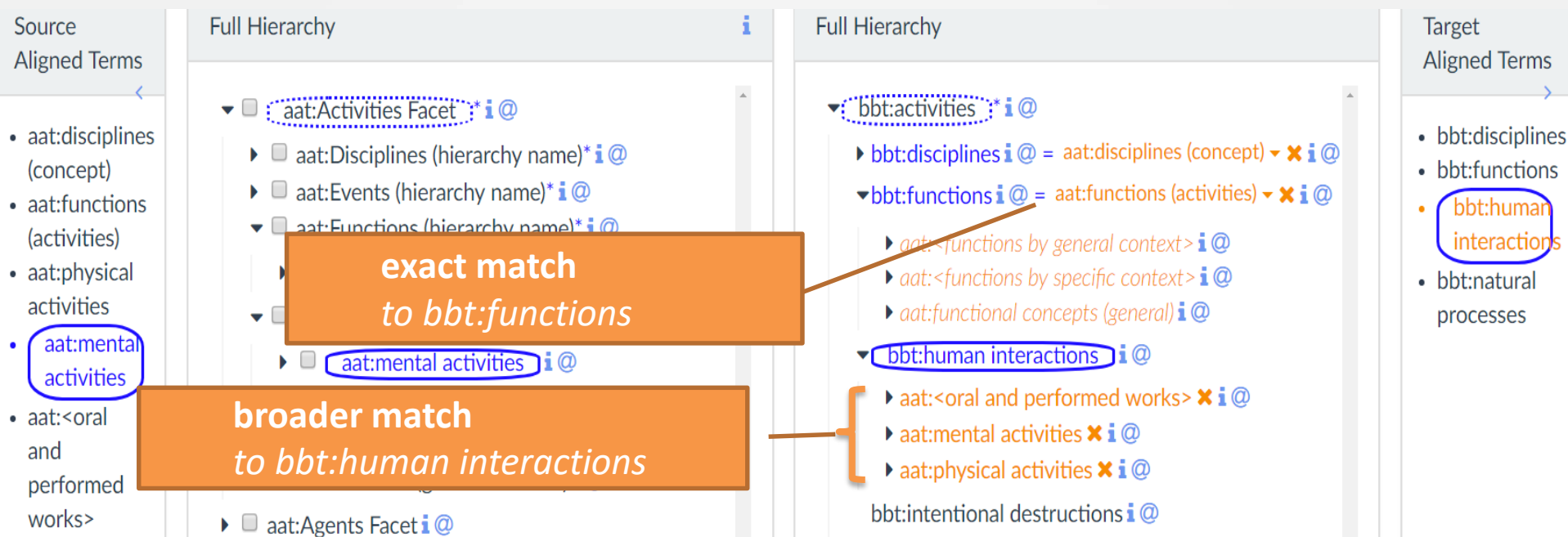
- ➔ - Directly aligned e.g. aat:mental activities, bbt:human interaction
- ➔ - Indirectly aligned e.g. aat:functional concepts (general)
- ✗ - Non-aligned e.g. bbt: intentional destructions

## VisTA Features 4/6

Source Aligned Terms	Full Hierarchy	Full Hierarchy	Target Aligned Terms
<ul style="list-style-type: none"> <li>• aat:disciplines (concept)</li> <li>• aat:functions (activities)</li> <li>• aat:physical activities</li> <li>• <b>aat:mental activities</b></li> <li>• aat:&lt;oral and performed works&gt;</li> </ul>	<p>Full Hierarchy</p> <ul style="list-style-type: none"> <li>▼ <b>aat:Activities Facet</b> *i@             <ul style="list-style-type: none"> <li>▶ <input type="checkbox"/> aat:Disciplines (hierarchy name)*i@</li> <li>▶ <input type="checkbox"/> aat:Events (hierarchy name)*i@</li> <li>▼ <input type="checkbox"/> aat:Functions (hierarchy name)*i@                 <ul style="list-style-type: none"> <li>▶ <input type="checkbox"/> aat:functions (activities)i@</li> </ul> </li> <li>▼ <b>aat:Physical and Mental Activities (hierarchy name)</b> <ul style="list-style-type: none"> <li>▶ <input type="checkbox"/> <b>aat:mental activities</b> i@</li> <li>▶ <input type="checkbox"/> aat:physical activities i@</li> </ul> </li> <li>▶ <input type="checkbox"/> aat:Processes and Techniques (hierarchy name)i@                 <ul style="list-style-type: none"> <li><input type="checkbox"/> aat:activities (general context)i@</li> </ul> </li> <li>▶ <input type="checkbox"/> aat:Agents Facet i@</li> </ul> </li> </ul>	<p>Full Hierarchy</p> <ul style="list-style-type: none"> <li>▼ <b>bbt:activities</b> *i@             <ul style="list-style-type: none"> <li>▶ bbt:disciplines i@ = aat:disciplines (concept) ▼ ❌ i@</li> <li>▼ bbt:functions i@ = aat:functions (activities) ▼ ❌ i@                 <ul style="list-style-type: none"> <li>▶ aat:&lt;functions by general context&gt; i@</li> <li>▶ aat:&lt;functions by specific context&gt; i@</li> <li>▶ aat:functional concepts (general) i@</li> </ul> </li> <li>▼ <b>bbt:human interactions</b> i@                 <ul style="list-style-type: none"> <li>▶ aat:&lt;oral and performed works&gt; ❌ i@</li> <li>▶ aat:mental activities ❌ i@</li> <li>▶ aat:physical activities ❌ i@</li> </ul> </li> <li>bbt:intentional destructions i@</li> </ul> </li> </ul>	<p>Target Aligned Terms</p> <ul style="list-style-type: none"> <li>• bbt:disciplines</li> <li>• bbt:functions</li> <li>• <b>bbt:human interactions</b></li> <li>• bbt:natural processes</li> </ul>

- **Visualization** of Correspondences
  - Highlight of the related terms at both hierarchies  
e.g. **aat:mental activities** and **bbt:human interactions**

## VisTA Features 5/6



- **Visualization** of matching relations
  - **exact-match**
    - represented with “=”
    - the source term shares its children with the target term
  - **broader-match**



## VisTA Features 6/6

- **Searching**
  - By term-id
  - Free-text
- **Visualization** of the search result
  - The search result is highlighted in the full hierarchy
  - The parent hierarchy of the searched term is highlighted too

The screenshot displays the VisTA interface with three main panels:

- Search:** Contains a search bar with the text "<materials by origin>".
- Source Aligned Terms:** Lists the search results, including "aat:disciplines (concept)".
- Full Hierarchy of aat:gvp:broaderGeneric:** Shows a hierarchical tree structure. The path from "aat:Materials Facet" down to "aat:<materials by origin>" is highlighted in blue, indicating the parent hierarchy and the specific search result.

At the bottom of the hierarchy panel, a note states: "Click on the info button next to each term to get more information about the term".



## Configuration of the tool

- **metaphacts** platform provides an **extensible template mechanism** based on modern technologies (HTML5, Web Components, Handlebars)
  - Easy configuration
    - SPARQL **query** templates
      - Displaying hierarchies
      - Searching hierarchies
    - RDF **properties** (types, hierarchy, alignment)  
*e.g. `rdf:type`, `skos:broader`, `gvp:broader`, `skos:exactMatch`, `skos:broaderMatch`*
    - RDF **classes** of concepts  
*e.g. `skos:Concept`, `gvp:Concept`, `gvp:Subject`, `gvp:Facet`, `gvp:Hierarchy`, `gvp:GuideTerm`*

## Conclusions

- Terminologies and thesauri, contrary to other ontologies can be handled in more **precise ways** regarding alignment in order to produce **accurate** and **exploitable** results.
- VisTA, provides an **interactive** solution for the **exact** terminology alignment problem, required in the context of data provisioning and aggregation processes.
- We propose VisTA not as a competitive but as a **complementary** supportive solution for the work pending the manual phase of an alignment procedure.
- This is a work in progress:
  - More features
  - Integrate new alignment constraints
  - Evaluation

End of presentation

Thanks for your attention!

- Questions or comments?