

The BackBone Thesaurus and thesaurus alignment.

E. Tsoulouha

Linked Conservation Data Consortium
Terminology Workshop
6-7 June 2019
Stanford University

Outline

- ❑ The BackBone Thesaurus:
goals, principles and motivation
- ❑ The structure of the BackBone Thesaurus:
motivation for its facets & hierarchies
- ❑ BBTalk: a tool to manage and display the content of the BackBone Thesaurus
- ❑ Thesaurus Alignment through VisTA, a tool for visual terminology alignment.

The BackBone Thesaurus (BBT)

- ❑ The **BackBone Thesaurus** is the research outcome of work undertaken by the *DARIAH Thesaurus Maintenance WG* in an effort to design and establish a coherent overarching thesaurus for the humanities, under which specialist thesauri and structured vocabularies used across scholarly communities can be aligned and form a thesaurus federation.
- ❑ It is constructed and maintained collectively by a distributed group of DARIAH affiliated organizations (AA, DAI, FORTH, FRANTIQ), and its content is available in 4 languages (English, French, German, Greek).

Building the BBT; goals

- ❑ BBT aims at enabling cross-disciplinary resource discovery while ensuring compatibility with thesauri that cover highly specific scientific domains and areas of knowledge in development.

- ❑ It favors a loose integration of multiple thesauri, by means of mapping them to a small set of high level terms (facets & hierarchies).
 - ❑ For resource discovery recall is more important than precision
 - ❑ Terms for discovery can be much coarser than for documentation
 - ❑ The more generic the concept, the higher the recall

- ❑ Aligned vocabularies retain their autonomy rather than being subordinated under an all-encompassing thesaurus. They are accessed through the global thesaurus.

Building the BBT; principles

- ❑ BBT facets and hierarchies were obtained through a **bottom-up analysis** of classificatory systems used by humanities researchers.
- ❑ **Relations btw BBT terms:** hierarchical isA
- ❑ **BBT assumes an open-world position:** facets and hierarchies within them do not exhaust the domain they classify.
 - ❑ Complements of terms or negative definitions are not permitted.
- ❑ **BBT is deliberately open to revision:** extending the domain of its application, will probably call for the definition of new hierarchies in the BBT, in an orderly fashion (**BBTalk**)

Motivation for the BBT

- ❑ BBT concepts heavily draw on CIDOC-CRM classes.
- ❑ However, efforts have been made to **reconcile** them with the AAT facets & high-level hierarchies, to some extent.
- ❑ Seemingly a top-down approach, BUT really a bottom-up process:
 - ❑ the definition of the CRM is an abstraction over museum documentation standards and practices.
 - ❑ to be of use to the humanities researchers, BBT structure conform to their terminological systems (to some extent).

The Structure of BBT

Activities (facet)

- activities (top term)
- - disciplines
- - functions
- - human interactions
- - intentional desctructions

Conceptual Objects (facet)

- conceptual objects (top term)
- - symbolic objects
- - propositional objects
- - methods
- - concepts

Geopolitical Units (facet)

- geopolitical units (top term)

Groups and Collectivities(facet)

- groups and collectivities(top term)

Material Things (facet)

- material things (top term)
- - built environment
- - mobile objects
- - physical features
- - structural parts of material objects

Materials (facet)

- materials (top term)

Natural Processes (facet)

- natural processes (top term)
- - geneses
- - natural destructions

Roles (facet)

- roles (top term)
- - offices
- - roles of interpersonal relations

Types of Epochs (facet)

- types of epochs (top term)

Materials [CRM E57 Material, AAT Materials]

Activities (facet)

- activities (top term)
- - disciplines
- - functions
- - human interactions
- - intentional desctructions

Conceptual Objects (facet)

- conceptual objects (top term)
- - symbolic objects
- - propositional objects
- - methods
- - concepts

Geopolitical Units (facet)

- geopolitical units (top term)

Groups and Collectivities(facet)

- groups and collectivities(top term)

Material Things (facet)

- material things (top term)
- - built environment
- - mobile objects
- - physical features
- - structural parts of material objects

Materials (facet)

- materials (top term)

Natural Processes (facet)

- natural processes (top term)
- - geneses
- - natural destructions

Roles (facet)

- roles (top term)
- - offices
- - roles of interpersonal relations

Types of Epochs (facet)

- types of epochs (top term)

CRM Perdurants and related AAT concepts in BBT

Activities (facet)

- activities (top term)
- - disciplines
- - functions
- - human interactions
- - intentional destructions

Conceptual Objects (facet)

- conceptual objects (top term)
- - symbolic objects
- - propositional objects
- - methods
- - concepts

Geopolitical Units (facet)

- geopolitical units (top term)

Groups and Collectivities(facet)

- groups and collectivities(top term)

Material Things (facet)

- material things (top term)
- - built environment
- - mobile objects
- - physical features
- - structural parts of material objects

Materials (facet)

- materials (top term)

Natural Processes (facet)

- natural processes (top term)
- - geneses
- - natural destructions

Roles (facet)

- roles (top term)
- - offices
- - roles of interpersonal relations

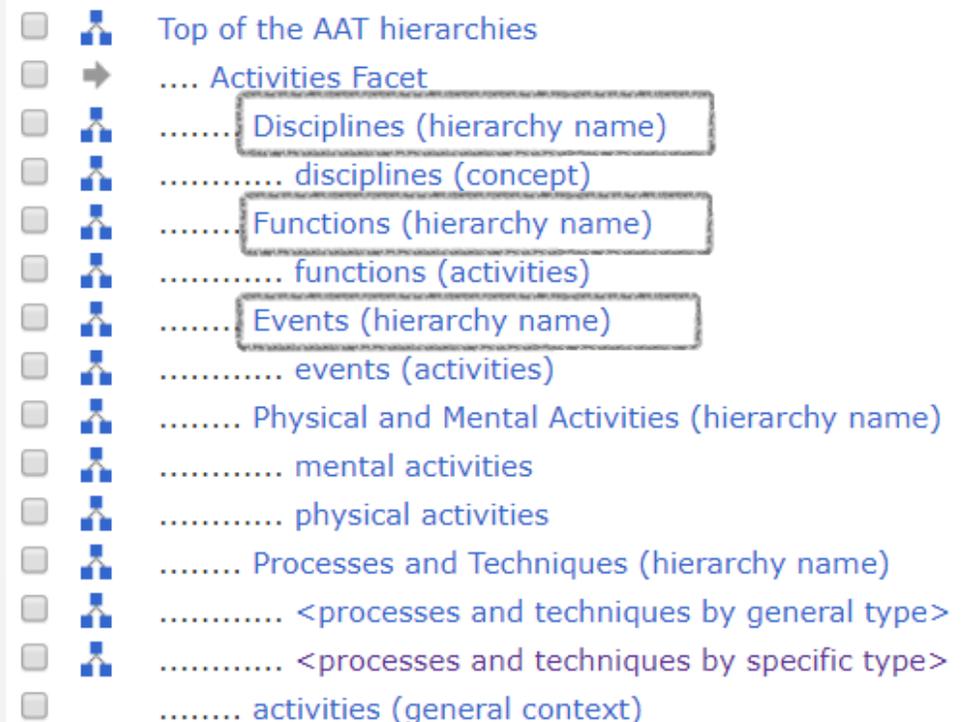
Types of Epochs (facet)

- types of epochs (top term)

- ❑ **BBT Activities facet:** a specification of the type of CRM **E7 Activity**
- ❑ Its most salient feature is **intentionality**.
- ❑ **It does not map to the AAT facet** –the latter does not presuppose intentionality.
- ❑ However, there has been an effort to partially maintain the structure of the AAT respective facet in the BBT hierarchies:

- ❑ disciplines
- ❑ functions
- ❑ human interactions

- ❑ **intentional destructions** on the other hand specify a subset of **E63 End of Existence** events, provided they were carried out by volitional agents.



- ❑ **Natural Processes** does not map to any CRM class in particular; rather, it is an abstraction over:
 - ❑ **geneses**:
 - ❑ the subset of **E63 Beginning of existence** events, characterized by the absence of a volitional agent.
 - ❑ **natural disasters**:
 - ❑ the subset of **E64 End of existence** events, characterized by the absence of a volitional agent.

- ❑ It does not map to the AAT either, where **natural events** isA **events** isA **activities**.

- ❑ **BBT Types of Epochs facet:** a specification of the type of CRM E4 Period.
- ❑ It differs from the AAT **Styles and Periods**, insofar as the latter also comprises instances of specific periods in Art and Architecture, whereas the BBT only permits **isA** relations between **types**.

- ❑  Top of the AAT hierarchies
- ❑  Styles and Periods Facet
- ❑  Styles and Periods (hierarchy name)
- ❑  <styles, periods, and cultures by general era>
- ❑  <styles, periods, and cultures by region>
- ❑  <generic styles, periods, and cultures>
- ❑  <styles, periods, and cultures by association>

CRM Perdurants and related AAT concepts in BBT

- ❑ Keeping the BBT as small as possible, means that some of the distinctions have to be made at a higher level.
- ❑ The BBT does not theorize on the properties of temporal entities, so there is no need to make the relation among the three facets explicit.

CRM Endurants & and related AAT concepts in BBT

Activities (facet)

- activities (top term)
- - disciplines
- - functions
- - human interactions
- - intentional desctructions

Conceptual Objects (facet)

- conceptual objects (top term)
- - symbolic objects
- - propositional objects
- - methods
- - concepts

Geopolitical Units (facet)

- geopolitical units (top term)

Groups and Collectivities(facet)

- groups and collectivities(top term)

Material Things (facet)

- material things (top term)
- - built environment
- - mobile objects
- - physical features
- - structural parts of material objects

Materials (facet)

- materials (top term)

Natural Processes (facet)

- natural processes (top term)
- - geneses
- - natural destructions

Roles (facet)

- *roles (top term)*
- - *offices*
- - *roles of interpersonal relations*

Types of Epochs (facet)

- types of epochs (top term)

Material Things facet

Material Things (facet):

- material things (top term)
- - built environment
- - mobile objects
- - physical features
- - structural parts of material objects

E18 Physical Thing

E25 Man Made Thing

E22 Man Made Object

E26 Physical Features

the physical things selected by **P46** is composed of

□ An area of almost complete overlap with the AAT.

.... Objects Facet

..... Built Environment (hierarchy name)

..... Components (hierarchy name)

..... Furnishings and Equipment (hierarchy name)

..... Object Genres (hierarchy name)

..... Object Groupings and Systems (hierarchy name)

..... Visual and Verbal Communication (hierarchy name)

Conceptual Objects facet

Conceptual Objects (facet)

- conceptual objects (top term)
- - symbolic objects
- - propositional objects
- - methods
- - concepts

E28 Conceptual Object

E90 Symbolic Object

E89 Propositional Object

E29 Design or Procedure

?

☐ Incompatibility with the AAT.

CRM Actors and related AAT concepts in BBT

- People that individually or in groups have the potential to perform actions intentionally, and be held responsible for them.
- Groups and collectivities** handles people acting collectively
- Actors perform Activities, in a given capacity and participate in other Events of specific type.
 - E7 Activity –P14 carried out by –E39 Actor
14.1 in the role of –E55 Type
Roles (facet)
 - E5 Event –P11 had participant –E39 Actor
roles of interpersonal relations
- The scope note definition of E74 Group foresees the extreme case of groups with exactly one member (f.i.: people holding offices)
 - offices**

CRM Actors and related AAT concepts in BBT

- ❑ **Geopolitical Units** represents a borderline case of Actor, one that combines organizational features (an administration and a jurisdiction) and spans over a geographical and temporal extent.
- ❑ Overall, there is significant overlap with the AAT Agents facet.
- ❑ AAT People map to BBT Roles and Groups & Collectivities (to the extent they do not exhibit organizational features),
- ❑ AAT Organizations (not administrative or political ones though) map to BBT Groups and Collectivities as well.
- ❑ The remainder of AAT Organizations map to BBT Geopolitical Units

Top of the AAT hierarchies

- Agents Facet
 - People (hierarchy name)
 - people (agents)
 - Organizations (hierarchy name)
 - organizations (groups)
 - Living Organisms (hierarchy name)
 - living organisms (entities)
 - biological components [N]
 - agents (general) [N]

BBTalk: A tool for federated thesaurus management

- ❑ BBT aims to bring together a federation of vocabularies spanning the Arts and Humanities and to render them interoperable.
- ❑ BBTalk is an online service for managing thesauri integration with the BBT and updating its content, when necessary.
- ❑ BBTalk offers a thesaurus alignment functionality allowing end users:
 - ❑ create links among specialist thesauri and the BBT .
 - ❑ provide evidence confirming/disconfirming the high-level terms of the thesaurus itself.
- ❑ <https://www.backbonethesaurus.eu/BBTalk/>

BBTalk: A tool for federated thesaurus management

End users can browse through the connections to BBT via BBTalk

Table of Local Thesauri Connections						
Connected Term	BBT Term	Submitter	Submission Date	Connection Relation	Connection Id	
Πετρώματα	materials	esavaidou	02.04.2019	Broader Match	2910	⚙️
Materialien	materials	Lena	18.12.2018	Exact Match	2071	⚙️
chemical element	materials	Tsoulouha	31.07.2018	Broader Match	2039	⚙️
Top Term: materials view mode				Broader Match	2038	⚙️
BBT Version: 1.2.2				Broader Match	2037	⚙️
Scope Note				Broader Match	2036	⚙️
This term classifies physical substances that are constituents of material objects or are used in their construction, but whose individual substance is not a factor in the objects' function (e.g. gold, water, bricks, etc.) The term may include pure raw materials, processed or unprocessed, and also materials that have been modelled and are incorporated into objects. Their main attribute is that they cannot be individualized, that is, they cannot be distinct "units" with clear and distinct boundaries or roles in respect to other units of the same or different kind.				Broader Match	2035	⚙️
				Broader Match	2034	⚙️
				Broader Match	2033	⚙️
				Broader Match	2032	⚙️
				Broader Match	2031	⚙️
Term Name Translations				Broader Match	2030	⚙️
Translation of Scope Note				Broader Match	2029	⚙️
URI				Broader Match	2028	⚙️
isomer				Broader Match	2027	⚙️
ion				Broader Match	2026	⚙️
insoluble substance				Broader Match	2025	⚙️
acid				Broader Match	2024	⚙️

Term Name Translations: υλικά Greek ▾

Translation of Scope Note: Ο όρος περιγράφει υλικές υποστάσεις που αποτελούν συστατικά υλικών αντικειμένων ή που χρησιμοποιούνται στην κατασκευή τους, χωρίς η ατομική τους υπόσταση να παίζει ρόλο στη λειτουργία του αντικειμένου (π.χ. χρυσός, νερό, τούβλα κ.ά.). Τα υλικά μπορεί να περιλαμβάνουν καθαρές πρώτες ύλες, επεξεργασμένες ή ανεπεξεργαστές, αλλά και υλικά που έχουν μορφοποιηθεί και ενσωματωθεί σε αντικείμενα. Χαρακτηριστικό γνώρισμα των υλικών είναι ότι δεν μπορούν να εξατομικευτούν, να αποτελέσουν δηλαδή διακριτές «μονάδες», με σαφή και διακριτά όρια ή ρόλο σε σχέση με άλλες μονάδες του ίδιου ή διαφορετικού είδους.

URI: <https://vocabs.dariah.eu/bbt/Concept/000003>

They can also navigate through the federated thesauri overarched by BBT and browse through their hierarchies, through the Federated Thesaurus Viewer.

<https://isl.ics.forth.gr/apollonis-federated-thesaurus/en/>

The screenshot displays the Skosmos web interface. At the top, the 'Skosmos' logo is on the left, and navigation links for 'Vocabularies', 'About', 'Feedback', and 'Help' are on the right. A search bar is located below the navigation links, with a dropdown menu set to 'from all' and 'English'. Below the search bar, the 'Skosmos Vocabulary Categories' section lists 'GENERAL CONCEPTS' and several thesauri: Backbone Thesaurus, DAI Thesaurus, DYAS HUMANITIES THESAURUS, GEMET, and Parthenos Vocabularies. The main content area is titled 'DYAS HUMANITIES THESAURUS' and includes a search bar with 'Content language' set to 'English'. On the left, there are tabs for 'Alphabetical', 'Hierarchy', and 'Groups'. The 'Alphabetical' tab is active, showing a list of concepts: '1626 Rocks' and '175 Marble'. The '1626 Rocks' concept is selected, and its details are shown on the right. The details include: 'PREFERRED TERM: 1626 Rocks', 'NARROWER CONCEPTS: 175 Marble', 'SCOPE NOTE: Naturally occurring materials of consolidated or unconsolidated mixture of minerals or/ and amorphous matter that make up the rigid outer layer of the Earth. They are classified into igneous, sedimentary and metamorphic rocks.', 'CONTRIBUTOR: esavaidou', 'CREATOR: esavaidou', 'IN OTHER LANGUAGES: Πετρώματα (Greek)', 'URI: https://humanitiesthesaurus.academyofathens.gr/dyas-resource/Concept/1626', 'Download this concept: RDF/XML TURTLE JSON-LD last modified 2019-04-01 Created 2019-03-29', and 'BROADER MATCHING CONCEPTS: 000003materials Backbone Thesaurus'.

BBTalk: A tool for federated thesaurus management

- ❑ Clashes between BBT concepts and thesauri aligned to it can be resolved by updating the content and/or structure of the BBT.
- ❑ Updating the BBT involves the introduction of new high-level terms plus the deletion/merging or splitting of existing ones.
 - ❑ BBTalk keeps track of such actions and allows for version control.
- ❑ Changes are implemented after careful consideration and discussion among the BBT managing team, until consensus is reached.
 - ❑ BBTalk supports discussions among end-users and its managing team.
- ❑ BBTalk supports multilingual content.

BBTalk: A tool for federated thesaurus management

Submission form for creating a new term or facet for the BBT

[view mo](#)

Submitter: Eleni Tsouloucha Institution: FORTH Field of Interest: Linguistics, Formal Semantics, Cognitive Semantics, Historical Semantics

Submission Date: 18.09.2018 Version: 1.2 Submission Id: 2037 Status: under discussion

New BBT Term or Facet name language [PARTHENOS]

Type Term

Scope Note This term classifies types of communications systems comprising of a finite set of elements and a set of recursive rules to combine them into a potentially infinite array of discrete expressions.

Broader Term conceptu

[Change Status](#)

[Forward to reviewer](#)



DISCUSSION (Visible to all curators and to reviewer: fhelfer, hgautier, MaurizioSanesi)

[View older comments](#)

Tsoulouha :

Date: 23.03.2019

Apparently BBT cannot accommodate sub-hierarchies at the moment. So the question remains: do we go for a generic "Language" hierarchy that is broad enough to serve as hook for both natural AND formal languages to link to? In this case, the distinction would be made at the level of micro-thesauri/specialist thesauri linked to BBT, not BBT itself.

Alternatively, do we take the similarities between natural and formal languages to be trivial/superficial and opt for two separate hierarchies within BBT?

this question is really open for debate i think.

as for artificial languages, like Esperanto. The fact that there are Esperanto speaking communities (whose native language is Esperanto) suggests that the boundaries btw artificial/constructed and natural language are very thin.

reference from wikipedia (<https://en.wikipedia.org/wiki/Esperanto>)

Finnish linguist Jouko Lindstedt, an expert on native-born Esperanto speakers, presented the following scheme[107] to show the overall proportions of language capabilities within the Esperanto community:

1,000 have Esperanto as their native language.

10,000 speak it fluently.

100,000 can use it actively.

One million understand a large amount passively.

Thesaurus Alignment

- ❑ The purpose of terminologies alignment is to integrate knowledge, to support the communication among individuals, teams and communities coming from different scientific, cultural, technical and other domains.
 - ❑ to integrate different terminologies under one target terminology used as a core terminology in a knowledge aggregator. The core terminology gets extended with new specialized terms.
 - ❑ to empower 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.

ViSTA: a Visual Terminology Alignment tool

ViSTA offers a simple and friendly web-based user interface for an **exact, target-driven, human-interactive** terminology alignment.

The alignment follows three principles:

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

VisTA key features

- The adaptation of the structure of the source terminology is supported by the tool: **removal of the incompliant subterms** that break the subsumption of the target

Full Hierarchy of <http://w3id.org/lob> skos:broader

- ▶ actors i@
- ▼ **attributes i@**
 - bypass pattern* i@
 - endleaf formats* i@
 - grain direction* i@
 - heat shrinkage temperature* i@
 - multi-section pattern* i@
 - ▶ relative size i@

Full Hierarchy of <http://vocabs.dariah.eu/bbt> skos:broader

- ▶ activities i@
- ▼ **conceptual objects i@**
 - concepts i@
 - ▼ **methods i@**
 - ▶ **attributes x i@**
 - ▶ **techniques x i@**
 - propositional objects i@
 - ▶ symbolic objects i@ = **symbolic objects** ▼ **x i@**

- ❑ Multiple-inheritance of terms is allowed (unless it breaks subsumption)
- ❑ The extension of the target terminology is based on broader match and the exact match relations

Full Hierarchy of <http://w3id.org/lob> skos:broader ▾

- ▶ actors i@
- ▶ attributes i@
- ▶ features i@
- ▶ materials i@
- ▶ objects i@
- ▶ places i@
- ▶ symbolic objects i@
- ▶ techniques i@

Full Hierarchy of <http://vocabs.dariah.eu/bbt> skos:broader ▾

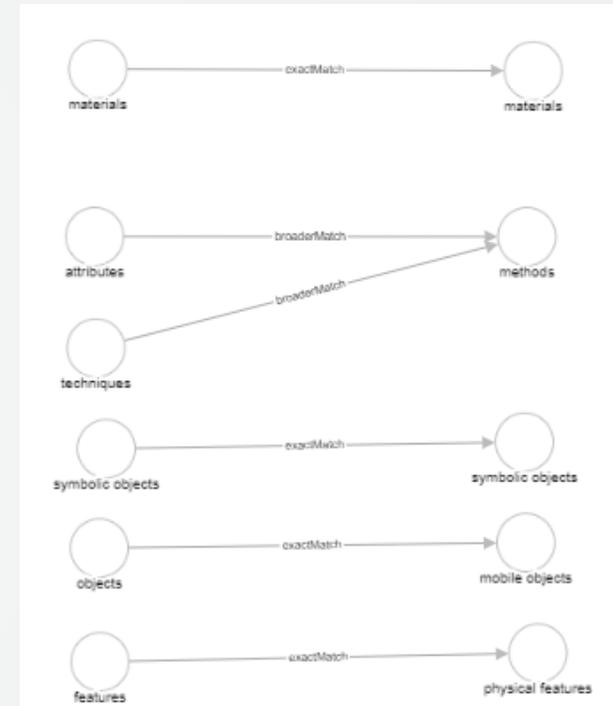
- ▶ activities i@
- ▶ conceptual objects* i@
- ▶ geopolitical units i@
- ▶ groups and collectivities i@
- ▶ material things* i@
- ▶ materials i@ = materials ▾ × i@
- ▶ natural processes i@
- ▼ roles i@
 - ▶ actors × i@
 - ▶ offices i@
 - ▶ roles of interpersonal relations i@
- ▶ types of epochs i@

The result of the Alignment

For each pair of aligned terminologies, ViSTA produces an alignment graph in RDF/SKOS.

The graph contains:

- ❑ the correspondences between terms
 - ❑ the children hierarchy of the source terms
- ❑ Searching space:
- ❑ Alignment graph + Target terminology graph



To align or not to align?

And more importantly what target to align to?

It really depends on one's purpose

BBT alignment:

- ❑ A fairly easy and quick process, given its small set of facets & hierarchies
- ❑ Terminologies are linked at a high-level to the BBT and preserve their autonomy
- ❑ One can publish their thesaurus and become part of the thesaurus federation
- ❑ Assuming there is no appropriate node with which to align, one can always propose a new branch within BBT (to be examined by the managing team)
- ❑ At the same time, (s)he can fully align a terminology or a subpart thereof to another terminology discovered through Skosmos (in the future)

Alignment with a fully-fledged thesaurus:

- ❑ It's a massive task, given the depth of the relations.
- ❑ It also addresses a different problem:
Wanting to reorganize a facet or other part of one's thesaurus and looking for a good terminology to align to!

The image displays two screenshots of a thesaurus interface, likely from a software tool like Thesauri Manager. The left screenshot shows a hierarchical tree structure for 'materials'. The right screenshot shows a similar structure for 'Materials Facet'.

Left Screenshot: materials*

- actors*
- attributes
- features
- materials*
 - adhesives*
 - animal materials*
 - covering material
 - decorative materials*
 - fibre materials*
 - gemstone
 - glass
 - metal
 - paper pulp
 - patched (materials)
 - pieced (materials)
 - plant materials
 - re-used (materials)
 - sewing thread
 - sheet (materials)*
 - strip materials*
 - surface coating materials
 - writing materials*

Right Screenshot: aat:Materials Facet*

- aat:Materials Facet* = materials*
- aat:Materials (hierarchy name)*
 - aat:materials (matter)*
 - aat:<materials by composition>*
 - aat:combination inorganic/organic material*
 - aat:atmosphere
 - aat:combination inorganic/organic animal material*
 - aat:<bone and bone components materials>*
 - aat:<bone components as materials>
 - aat:bone (material)* = bone*
 - aat:<shell and shell material>*
 - aat:mother of pearl* = mother of pearl*
 - aat:shell (animal material)
 - aat:coral (material)
 - aat:pearl (animal material)
 - aat:composite material
 - aat:organometallic compound
 - aat:soil
 - aat:inorganic material*
 - aat:alkylalkoxysilane

One can't be sloppy!!

Full Hierarchy of <http://w3id.org/lob> skos:broader

- almanacs i
- bindings i
- codex-form books i
- components i
 - attachment components i
 - board attachment components i
 - board components i
 - boards i
 - bookblock components i
 - bookblocks i
 - bookmark components i
 - bookmarks i
 - cover components i
 - covers i
 - enclosure covers i
 - endband components i
 - endbands i
 - envelope flap components i
 - fastening components i
 - fastenings i
 - furniture (components) i
 - labels i

Full Hierarchy of [aat](#) gvp:broaderGeneric

- aat:Built Environment (hierarchy name) i
- aat:Components (hierarchy name) i
 - aat:components (objects parts) i = components - x i
 - aat:<components by general context> i
 - aat:<components by specific context> i
 - aat:<tool and equipment components> i
 - aat:ammunition components i
 - aat:architectural elements i
 - aat:biological components i
 - aat:building divisions i
 - aat:container components i
 - aat:costume components i
 - aat:furnishings components i
 - aat:information form components i
 - aat:material components i
 - aat:measuring device components i
 - aat:sound device components i
 - aat:system components i
 - aat:textile components i
 - aat:vehicle components i
 - aat:visual works components i
 - aat:weapon components i
 - attachment components x i
 - board attachment components x i
 - board components x i
 - boards x i
 - bookblock components x i
 - bookblocks x i
 - bookmark components x i
 - bookmarks x i
 - cover components x i
 - covers x i
 - enclosure covers x i

Rather, (s)he must establish exact correspondences, which requires an in-depth knowledge of both **source** and **target** vocabularies

