

Controlled Vocabularies for seismology

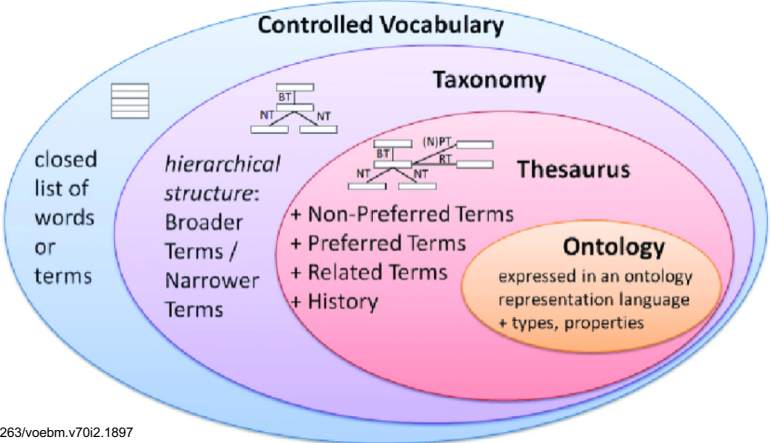
Angelo Strollo, Florian Haslinger
and ...

Controlled Vocabulary: also called an authority file, is an authoritative list of terms to be used in indexing (human or automated). A controlled vocabulary for a project might actually include multiple authority files for different kinds of terms. Controlled vocabularies do not necessarily have any structure or relationships between terms within the list. Controlled vocabularies are often used for name authorities (proper nouns), such as persons, organization names, company names, etc. Controlled vocabularies are the broadest category, which includes thesauri and taxonomies. Thesauri and taxonomies are specific kinds of controlled vocabularies, but not all controlled vocabularies are thesauri or taxonomies.

Taxonomy: typically a controlled vocabulary with a hierarchical structure, with the understanding that there are different definitions of a hierarchy. Terms within a taxonomy have relations to other terms within the taxonomy. These are typically: parent/broader term, child/narrower term, or often both if the term is at mid-level within a hierarchy.

Thesaurus: as used in information science and literature retrieval, is essentially a controlled vocabulary following a standard structure, where all terms in the thesaurus have relationships to each other. These relationships are typically of three kinds: hierarchical (broader term/narrower term), associative (see also), and equivalent (use/used from or see/seen from). In addition, it is common in thesauri that some or all terms have scope notes, brief explanations of how the term should be used in indexing. Term history notes may also be present.

An **ontology**, like a thesaurus, is a kind of taxonomy with structure and specific types of relationships between terms. In an ontology the types of relationships are greater in number and more specific in their function. Relationships could include, for example, located in to relate an organization to a place, produces/is produced by to relate a company and its product, and employs/employed by to relate a company and a person. Information, that in a simple controlled vocabulary or taxonomy is conveyed through indexing, is embedded into the ontology itself.



Motivation within the seismological community and beyond

- Almost no usage of controlled vocabularies in seismology
 - NASA, NERC, EPOS started started a task force on this topic some years ago
- Need of expanding metadata descriptions with key words for non seismologists
- Facilitate data discovery by tagging datasets (when possible with key words)
- Increasing usage of multidisciplinary data: AI applications, interoperability becoming a reality for some disciplines => users of seismological services that are not seimologists
- Funding agencies demanding integration and interoperability towards the cloud services
- Running projects (Geo-INQUIRE, DT-GEO, SUBMERSE, ... others in Europe) attempting to integrate more exotic data types and products in the well established seismological archives

The first step now at this meeting

Is this topic considered relevant for this group?

If YES, include this as an action item for the near future, the WG will start working on this, possible initial actions:

- Identify a group of people/institutions with experience on this topic or willing to work on it
- Survey within this group on the existing situation and use cases
- Merge the existing list of terms collected by the group
- Presents the results back to the WG and discuss next steps which may include interaction with other associations