Controlled Vocabularies for seismology (2023)

Angelo Strollo, Florian Haslinger and …

2025 Reload!

~~Controlled Vocabularies for seismology~~

Tags* for datasets in seismology

Angelo Strollo, Florian Haslinger and ?

*tags (keywords or metadata) are used to categorize, organize, and search for information within the dataset. Tagging is a method to add descriptive labels to datasets, often containing information like the author, date, department, or content type, enabling users to efficiently filter, discover, and manage large volumes of data.

Possible next steps (to be discussed in the WG) **2023**

Is this topic considered relevant for this group?

If YES, how to structure the discussion?

- Identify some common use cases and start the discussion within this WG
  - OBS and Mobile Pools data
  - DAS products (marine and land applications)
  - Environmental data
  - GNSS products
  - Other datasets in archive from Marine, Volcanological, Near-fault Observatories
- Collect existing list of words or terms with hierarchy (Taxonomy)
- Create an initial merged list of terms (reviewing what the groups have already with respect to existing high level controlled vocabularies)
- Start with taxonomy events type separated from data type

Possible next steps (to be discussed in the WG) Reloaded!

1. Identify existing sources (glossaries, vocabularies) to which we usually refer to
    a. SEED manual https://www.fdsn.org/pdf/SEEDManual_V2.4.pdf
    b. StationXML documentiaion https://docs.fdsn.org/projects/stationxml/en/latest/
    c. QuakeML documentation https://quake.ethz.ch/quakeml/Documents
    d. NMSOP2 glossary https://bib.telegrafenberg.de/publizieren/bibliotheksverlag/nmsop
    e. SeisComP glossary https://www.seiscomp.de/doc/base/glossary.html
    f. USGS glossary https://www.usgs.gov/glossary/earthquake-hazards-program
    g. ISC event type nomenclature https://www.isc.ac.uk/standards/event_types/event_types.pdf
    h. EPOS-Seismology vocabulary https://registry.epos-eu.org/ncl/FAIR-Incubator/_tcs-SEISMO
    i. Any other source to suggest? Kept out for the moment large vocabularies in use already
2. Organize the sources in a Taxonomy with AI tools, Natural Language Processing (NLP) and clustering algorithms
3. Expert review to validate and refine the AI's output (anyone here willing to help?)

Taxonomy for Seismology with AI tools

1. Collect Content and identify Keywords
    a. gather the text or data to organize, such as documents, descriptions, glossaries, etc.
    b. Extract key terms and concepts from the content, as these will form the basis of the taxonomy.
2. Apply NLP and Clustering, leverage LLMs for Suggestions
    a. AI tools use NLP to understand the meaning of the content and clustering algorithms to group related items based on their semantic connections.
    b. Large Language Models (LLMs) can suggest related concepts, definitions, and hierarchical relationships to expand the taxonomy, eliminating the need to create terms from scratch.
3. Refine and Validate the Taxonomy
    a. Domain experts should review the AI-generated categories and terms to ensure they accurately reflect the domain.
    b. Use AI-powered tools to check for consistency, ensuring terms are used uniformly across the taxonomy and align with established definitions.