



## **Bridging fragmented terminologies: advancing vocabulary harmonization in Seismology through AI and community co-creation**

**Juliano Ramanantsoa**<sup>1</sup>, Angelo Strollo<sup>2</sup>, Florian Haslinger<sup>3</sup>, Javier Quinteros<sup>2</sup>, Daniele Bailo<sup>4</sup>, Otto Lange<sup>5</sup>, Samshuijzen Laurens<sup>5</sup>, Sven Peter Naesholm<sup>6</sup>, and Mathilde B. Sørensen<sup>1</sup>

<sup>1</sup>Department of Earth Science, University of Bergen, Norway

<sup>2</sup>GEOFON, Section 2.4, GFZ Helmholtz Centre for Geosciences

<sup>3</sup>Swiss Seismological Service, ETH Zurich

<sup>4</sup>Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

<sup>5</sup>University Library, Utrecht University, Netherlands

<sup>6</sup>NORSAR, Kjeller, Norway

The conceptual clarity of any scientific field depends fundamentally on the precision and standardisation of its terminology. Prior studies have shown that an absence of standardized terminologies can lead to interpretive ambiguity, imprecise outputs, and divergent interpretations across research communities. In seismology, terminologies remain scattered across institutional glossaries, impeding data FAIRness (Findability, Accessibility, Interoperability, and Reusability), metadata consistency, and collaboration with adjacent fields such as transdisciplinary research and AI engineering.

This work, carried out within the Geo-INQUIRE\* project, introduces a vocabulary generation framework and a prototype database implementing three integrated innovations that consolidate the sparse seismological terminologies into a structured, machine-readable format: i) authority-first retrieval, ii) AI-mediated semantic triangulation, and iii) participatory expert governance.

The authority-first pathway performs weighted, priority-ranked extraction from eight expert-curated data centre sources (including FDSN, USGS, EarthScope, EPOS, and other relevant documents from the community), ensuring that the definitions originate from trusted references. The AI fallback pathway is activated only when authoritative retrieval fails, employing a semantic triangulation method in which three large language models - such as OpenAI's GPT-5.2, Anthropic's Claude Opus 4.5, and Google's Gemini 3 - independently generate candidate definitions. Embedding-based similarity analysis determines synthesis eligibility; if cross-model agreement falls below 50 percent, an expert flag is raised to prevent semantic uncertainty. When synthesis proceeds, a transparent concept-merging process extracts common and unique contributions from each model, recording all reasoning steps and preserving full provenance, overcoming a critical limitation of black-box AI knowledge generation.

Beyond technical generation, this work embeds vocabulary development within a participatory

framework that transforms terminology from static definitions into community-validated knowledge. Through structured digital deliberation involving more than ten domain experts via a GitHub-based workflow, the approach delivers transparency, auditability, and collective ownership. Experts validate AI-retrieved content, resolve edge cases, and steward terminology evolution through documented discussion threads, ensuring definitions reflect both institutional authority and practitioner consensus while fostering public trust in seismology.

The system produces vocabulary encoding scheme-compliant entries with dual definitions: an authoritative version weighted by source priority, and an AI-synthesized alternative with full provenance. The source-weighting mechanism is fully flexible ensuring the reusability of the framework. Applied to over 500 terms across 4 thematic clusters, this framework demonstrates that AI can systematically extend vocabulary completeness while participatory governance safeguards epistemic integrity. By coupling algorithmic precision with community oversight, this framework strengthens data discovery, metadata coherence, and research infrastructure interoperability across European and international seismological networks that advance transparent, reproducible, and interoperable seismological science.

\*Geo-INQUIRE (Geosphere INfrastructures for QUEStions into Integrated REsearch) is funded by the European Union (GA 101058518).