EGU24-18951, updated on 11 Mar 2024
https://doi.org/10.5194/egusphere-egu24-18951
EGU General Assembly 2024

Adopting FAIR data management practices in mountain hazard research: Strategies for ensuring data quality for landslide susceptibility modeling

Laura Waltersdorfer[1], Andrea Siposova[1], Matthias Schlögl[2], and Rudolf Mayer[1]
[1]SBA Research, Austria
[2]GeoSphere Austria

Mountainous regions such as the Austrian Alps face a constant threat of natural hazards. Over time, this persistent danger has prompted a transition from heuristic hazard management strategies towards a more quantified risk culture. Since quantitative risk assessment heavily relies on understanding the occurrence frequency of the hazard processes under consideration, knowledge about past events and their characteristics becomes pivotal, thereby shaping the effectiveness and broader applicability of methodological workflows employed in this context.

We present challenges, and insights gleaned from the research project “gAia”, focusing on a data-driven susceptibility assessment for shallow landslides in Austria. The identified challenges mainly revolve around the quality of landslide inventories, which is influenced by factors like underreporting, inconsistent documentation, and lack of standardized data management practices. We thus recommend adopting FAIR (Findability, Accessibility, Interoperability, Reusability) principles and developing Data Management Plans to address these issues, and propose a general data management workflow:

- Identify data sources and contents: Collect information about data sources and characteristics in a (machine-readable) DMP to obtain an overview of all data sources and most important characteristics (e.g. format, size, license, context, bias limitations). This should support the contextualization and ability to reuse this data.
- Define processing activities: Explicitly define processing workflows to enhance reproducibility and transparency, using established standards such as Business Process Management (BPMN) or semantic web technologies to represent complex processes formally and make them more comparable and accessible to users.
- Define (meta)-data and process activities trace templates: Provide metadata templates for datasets and trace processing activities to improve interoperability and reusability. Define domain-specific vocabularies and use concepts such as datasheets, model cards, ML experiment tracking and model registry tools as well as task orchestration platforms for data engineering pipelines to make results more traceable and reviewable.
- Monitoring processes for natural hazard event data: Implement processes to ensure adherence to quality metrics, with results published in machine-readable formats.

We detail the implementation of these steps using established concepts of traceability and provenance, and encourage to implement workflow tasks using common open source programming languages. In addition, we endorse the use of Git for version control and GitLab/GitHub as tools for facilitating collaboration and structuring technical tasks.

The benefits of the proposed data management strategies for enhancing quality and reliability of data as well as increasing overall transparency of processes are showcased in the gAia project. The project workflow, represented as a P-Plan, demonstrates the application of these strategies in different phases. Specifically, the importance of proper data management and adherence to FAIR principles for data-driven research and practical usability is highlighted using landslide inventories as a core example.

In summary, we provide insights into the complexities of geospatial data management in mountain hazard research and offer practical solutions to enhance the integrity and reliability of data for supporting effective risk assessment and disaster risk reduction.

gAia is funded through the KIRAS Security Research Program for Cooperative Research and Innovation Projects by the Austrian Research Promotion Agency (FFG) and the Federal Ministry of Finance, under grant agreement FO99988636910.