Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.35987

Geochemical Big Data for Quantitative Studies of Magmatic Systems: a Perspective from global compilation databases

LEANDER KALLAS[1], MARIE KATRINE TRAUN[1], MARTHE KLÖCKING[2], BÄRBEL SARBAS[1], PENG JI[3], KERSTIN LEHNERT[4], MATTHIAS WILLBOLD[1] AND GERHARD WÖRNER[1]

[1]University of Göttingen
[2]University of Münster
[3]Lamont-Doherty Earth Observatory of Columbia University
[4]Columbia University

Data-driven approaches are transforming the study of magmatic systems by enabling quantitative analysis of geochemical variability across tectonic settings and geological timescales. They rely on large, curated, interoperable datasets that serve as inputs for statistical analyses, thermodynamic models, and machine-learning workflows. Developed over more than 25 years and now maintained within the Digital Geochemical Data Infrastructure (DIGIS) project, the GEOROC database (Geochemistry of Rocks of the Oceans and Continents) is one of the most comprehensive global compilations of igneous and metamorphic rock and mineral compositions. Contributing to the EarthChem Portal, GEOROC supports a federated ecosystem delivering more than 50 million geochemical values from over 30,000 publications to the community. These efforts also support the OneGeochemistry initiative, which promotes global standards and best practices for FAIR (Findable, Accessible, Interoperable, and Reusable) geochemical data.

GEOROC- and EarthChem-enabled data compilations support data-driven investigations across a wide range of magmatic research areas. Machine-learning approaches applied to such datasets can discriminate tectono-magmatic settings, identify global petrogenetic processes, and predict magma storage and evolution. Other applications include modelling mineral formation conditions, assessing magma-crust interactions, constraining crustal growth and long-term compositional variability, and defining the type and origin of mantle magma sources across tectonic settings. Large datasets thus enable reproducible benchmarking of petrological and geochemical data-driven models and provide a foundation for next-generation computational studies of magmatic systems across diverse tectonic and volcanic regimes.

As computational methods evolve, interoperable, high-quality datasets increasingly become a limiting factor for scientific progress. We will present examples of how global, machine-readable, and quality-controlled geochemical data and metadata from GEOROC and EarthChem, enable interoperable and high-quality data compilations and closer alignment between database curation and the needs of computational magmatic researchers. These compilations are supported by controlled vocabularies, harmonized data structures, rigorous metadata evaluation, and active community feedback. Combined with computational tools, these resources open new opportunities to resolve multiscale magmatic processes and compare magmatism across environments. We will also draw attention to the need for consistent data structures and metadata standards in geochemistry that currently hinder more automated ingestion of new data into the global databases.