Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.27458

The Future of GEOROC: Database Curation for the Geochemical Community

MARIE KATRINE TRAUN AND TEAM DIGITAL GEOCHEMICAL INFRASTRUCTURE FOR GEOROC (DIGIS)

Georg-August-Universität Göttingen

The GEOROC database (Geochemistry of Rocks of the Oceans and Continents) was born in 1999 at the Max Planck Institute in Mainz (Germany) under the leadership of A. W. Hofmann and B. Sarbas. For 25 years, GEOROC has served the geochemical community. Since 2021 the database has been curated and further developed from its new home at the University of Göttingen. Historically, geochemical data acquisition has been done by local studies and research groups. The lack of common data standards in method and acquisition reporting (metadata, vocabularies, etc.) and data formats did not allow for easy data accessibility and reuse. As the amount of published geochemical data has grown, making global compilations of well-curated geochemical data with harmonized metadata have become an urgent task. This is the core mission of GEOROC. GEOROC now has more than 685000+ records for whole rock, glass and mineral analyses. During the past four years, the GEOROC IT infrastructure was modernized as part of the WorldFAIR funded OneGeochemistry Initiative and GEOROCs interoperability with PetDB (EarthChem) reestablished. GEOROC now also offers a data repository service where authors can publish datasets with citable DOIs. However, in the era of AI chatbots and scraping algorithms, data including geochemical tables can also be scraped from the internet into uncurated piles and fed to generalized data analysis models. In this context, the greatest strength of our well-curated data products is our expert reviewed data standard conventions, vocabularies, acquisition and metadata curation. Despite our best efforts, our community has given us very clear feedback: We need to do better, specifically with respect to data and metadata quality. Therefore, we are directing a new focus on developing data quality assessment methods in our database. These utilize statistical analysis, Machine Learning methods, and AI to aid in text analysis and metadata acquisition. Additionally, we are collaborating with our sister database GeoReM (Geological Reference Materials database) on cross-correlating reference materials and metadata between the databases in a NDFI4Earth funded project. How can we do better for our community and continue to make GEOROC a database for the future of geochemical research?