Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.31587

Streamlining OIB data compilation and processing: a Python-based workflow for GEOROC

MARTHE KLÖCKING AND ZACHARY T ERIKSEN
Universität Münster

Ocean island basalts (OIBs) are indirect samples of the Earth's interior that are invaluable for understanding the evolution of our planet. Their geochemical compositions provide unique insight into the chemical and thermal heterogeneity of the mantle. Global compilations that harmonise previously published data are the basis of a growing number of high-impact scientific studies. However, the effort required to ensure consistency and data quality is substantial, so these projects are often one-time initiatives that quickly become outdated and lose relevance. Here, we present a new Python-based workflow for a reproducible data query of the GEOROC database, followed by objective data filtering and quality assessment. This approach generates a consistent, high-quality compilation dataset that is easy to replicate and update, providing a reliable and comparable framework for future studies.

The GEOROC database has been compiling published OIB compositions since the late 1990s, and now represents a comprehensive collection of ocean island chemistry that is continuously being updated. Available OIB data have more than doubled since the first published compilations, with a current volume of around 80,000 samples from over 3,000 peer-reviewed journal articles. GEOROC provides free access to their data holdings and offer a diverse range of customisable query options thanks to extensive sample, geographical and analytical metadata. We present a modular Jupyter Notebook to demonstrate: (1) data querying and download; (2) filtering, cleaning and quality assessment; (3) normalization and transformation; and (4) statistical data analysis. GEOROC's API enables command-line access to thousands of analyses using a diverse range of filters. This new programmatic interface allows for fast, flexible and reproducible data queries that overcome many of the limitations of the existing graphical user interface. Our workflow demonstrates the API usage with flexible query options, before working with the resultant dataset using common data analysis/manipulation tools available in the Python programming language. Additionally, we include optional modules, such as olivine fractionation correction and log-ratio transformation, to prepare compositional data for statistical analysis. By streamlining data compilation and processing, we hope this workflow will encourage more global-scale geochemical studies by enabling researchers to focus on scientific questions rather than data wrangling.