Goldschmidt 2024 Abstract
https://doi.org/10.46247/gold2024.24646

Geochemical Databases: Promise, Problems, and Prospects

CHARLES H. LANGMUIR[1], KERSTIN A LEHNERT[2] AND ROGER FU[1]

[1]Harvard University
[2]Columbia University
Presenting Author: langmuir@eps.harvard.edu

PetDB was a response to an RFP from the RIDGE program. At that time many geochemists thought a database was a one time effort leading to a spreadsheet. Gradually we realized the complexities of establishing a relational database, and that databases were living entities requiring constant attention. PetDB has become an essential resource for anyone working on oceanic basalts. Integration with GeoMapAp was an essential added capability that allowed integration with bathymetric and geophysical databases.

While databases have been a boon for global researchers, they can be abused by users, many of whom simply download data without the essential steps of evaluation and screening. Databases cannot serve as gatekeepers and have to include everything that is published. Published data is of variable quality, and there are inevitably errors. Surprisingly large interlaboratory variations create unnecessary noise. Careful screening and evaluation of the data is time-consuming—the MORB compilation of Gale et al. (2013) involved an entire Ph. D thesis, as did the dataset for convergent margins of Steve Turner. To help with the problems of bulk downloads (a “data dump”) both PetDB and GEOROC create curated datasets, but these often lack necessary geological context. One of the great applications of GEOROC is to investigate Earth history, but there the problems are even more daunting. In contrast to fresh young rocks from a known tectonic setting, ancient rocks are almost never fresh and have tectonic settings that are inevitably uncertain. Misleading “statistics” are often applied, such as binning inherently diverse global data and asserting a simple standard error to reflect the uncertainty of the mean, despite potential issues with highly correlated and under-sampled data points. Comparison to the complexity and diversity of the modern Earth provides important perspectives on such an approach.

To be most useful for the future, databases will need to sustain inclusiveness while providing properly curated datasets that can be downloaded with confidence. There are needs for systematic error correction to steadily improve database quality, a community-wide effort of interlaboratory calibration, and integration with sample repositories to make invaluable physical samples available for posterity.