

Providing analysis-ready data for computational geochemical modeling through geochemical data syntheses

LUCY PROFETA¹, KERSTIN A LEHNERT¹, GERHARD WÖRNER², MARTHE KLÖCKING², STEPHEN RICHARD¹, MR. LEANDER KALLAS, M.SC.², MANJA LUZI-HELBING³, PENG JI¹, MATTHIAS WILLBOLD², ADRIAN STURM², STEFAN MÖLLER-MCNETT², JUAN DAVID FIGUEROA¹, SEAN CAO¹, BÄRBEL SARBAS², ANNIKA JOHANSSON¹ AND HANNAH A SWEETS¹

¹Columbia University

²Georg-August-Universität Göttingen

³Geo.X - Research Network for Geosciences in Berlin and Potsdam / GFZ German Research Centre for Geosciences

Presenting Author: lprofeta@ldeo.columbia.edu

With more than two decades of careful data curation and innovative cyberinfrastructure use, data syntheses such as PetDB and GEOROC have been consistently providing FAIR and open analysis-ready data to the scientific community.

Over the past two years, PetDB and GEOROC have reaffirmed their commitment to working in tandem for the betterment of data availability for the geochemical community. With a signed memorandum of understanding, both teams have allocated resources to harmonizing technologies, vocabularies, and curation protocols. Data curation for new literature data has been divided between the GEOROC and PetDB teams to ensure full coverage of themes and locations associated with newly released geochemical data, as well as to avoid duplication of efforts. Data is also being harmonized by using IGSNs to identify data from the same analyzed samples. The results of these joint efforts are visible through the availability of more than 50 million data points through the EarthChem Portal [1]. Moreover, the development of APIs for both systems provides a pathway for integrating these data directly and seamlessly into modeling code.

These harmonization efforts also brought to light the need for joint vocabularies for seamless communication between the two systems. Wherever possible, the joint vocabularies used machine-readable external vocabulary standards and community best practices, such as the IUPAC Gold Book [2] or OpenMindat [3], supplemented by internal lists published by our systems, e.g. Analytical Methods for Geochemistry and Cosmochemistry [4].

The improvements derived from this partnership between GEOROC and PetDB to machine-readable, analysis-ready data availability will be instrumental to the computational geochemical modeling community, reducing the time needed for QA/QC and data harmonization that otherwise can delay or hinder such research endeavors completely.

[1] <http://portal.earthchem.org/>

[2] Chalk, Stuart, and Leah McEwen. "The iupac gold book: An exemplar for iupac asset digitization." *Chemistry international* 39.3 (2017): 25-30.

[3] Ma, Xiaogang, et al. "OpenMindat: Open and FAIR mineralogy data from the Mindat database." *Geoscience Data Journal* 11.1 (2024): 94-104.

[4] Lehnert, K. 2023. Analytical Methods for Geochemistry and Cosmochemistry. Concept Scheme for Analysis Methods in Geo- and Cosmochemistry. Published at Research Vocabularies Australia. Version 1.2. <https://vocabs.ardc.edu.au/viewById/650>.