# How Can We Sustain Geochemical Databases into the Future?

KERSTIN LEHNERT, LESLEY WYBORN[123]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA; lehnert@ldeo.columbia.edu
[2]Australian National University, Canberra, Asutralia

The geochemistry community today can look back on over two decades of online access to large-scale, global data geochemical databases including PetDB, EarthChem, and GEOROC that have inspired and supported hundreds of scientific studies, generating new knowledge and insights, and even leading to entirely new paradigms in research such as 'statistical geochemistry' that were not possible in times when individual researchers needed to spend years to compile their comprehensive syntheses of literature data. Geochemical databases provide easy access to millions of geochemical measurements that are accompanied by comprehensive and harmonized metadata for context and provenance to search, filter, sort, and evaluate the data. Web-based software tools allow users to search the databases for specific samples and specific measurements performed on them, and combine data from previously disparate sources into new, customized datasets. Citation records (the PetDB database now has >800 citations in the literature) demonstrate the broad use of the databases.

The databases were assembled by data managers, who extract data and metadata from text, tables, and supplements of publications for inclusion in the databases, and they are still maintained this way. While this procedure ensures the high quality of metadata, it is a time-consuming task due to the multitude of data formats, units, normalizations, vocabularies, etc., i.e. lack of best practices for geochemical data reporting, which means it is costly and therefore challenging to support and sustain. Long-term sustainability of databases such as PetDB, NAVDAT, and GEOROC is a problem that the geochemistry community needs to address jointly and on a global scale to ensure future availability of what has become a fundamental research infrastructure.

This presentation will explore opportunities presented by emerging international collaborations that engage relevant communities and stakeholders, including researchers, societies, funders, publishers, and data infrastructure providers in the Earth sciences in defining and planning steps toward the development of a sustainable global geochemical data network that will ensure preservation, discovery, and access of geochemical data through coordination and collaboration among international geochemical data providers.