Geochemical Databases: understanding their history, design, policies and principles to shape future global networks

MARTHE KLÖCKING[1], KERSTIN A LEHNERT[2] AND LESLEY A. I. WYBORN[3]

[1]University of Münster
[2]Lamont-Doherty Earth Observatory, Columbia University
[3]Australian National University
Presenting Author: mkloecking@uni-muenster.de

Geochemistry as a discipline centers around data. Modern laboratories produce large volumes of highly diverse data types that challenge established practices and capabilities for organizing, analyzing, preserving, and accessing these data. At the same time, sophisticated computational techniques, including machine learning, are increasingly applied to geochemical research questions, which requires easy access to large volumes of high-quality, well-organized and standardized data.

Data management has always been important in geochemistry but has recently become a necessity for the discipline to continue to thrive in the age of digitalization and artificial intelligence. This new treatise chapter on 'Geochemical Databases' summarizes the landscape of geochemical databases, including the different types of data systems based on their purpose, and their evolution in a historic context. We apply the life cycle model of geochemical data; explain how current standards, practices, and policies determine the design of modern geochemistry databases and data management; discuss the ethics of data reuse such as data ownership, data attribution, and data citation; and create a vision for the future of geochemical databases: data being born digital, connected to agreed community standards, and contributing to global democratization of geochemical data.