Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.29701

Enhancing Data Quality Through Expert Efforts: From Cyberinfrastructures to Cyberinfrastructures

DR. CHUNTAO LIU, ZHOU ZHANG AND Z.J.U. EARTH DATA TEAM

Zhejiang University

The rapid advancement of statistical methods and machine learning algorithms applied to data from cyberinfrastructures offers significant opportunities to uncover the secular evolution and chemistry of the solid Earth. However, these cyberinfrastructures, in their current state, contain raw data with missing categories, errors (including in age information), and misplaced chemical compositions that may conflict with published data. These inaccuracies primarily arise from the absence of standardized practices for publishing rock geochemistry data and limitations in optical character recognition techniques used to convert tablet data into readable formats. As a result, the data quality issues in cyberinfrastructures are not inherent to the infrastructure itself, but rather to the data publishing format.

Before the data input process to cyberinfrastructures can be fully standardized to ensure high-quality data, an alternative approach involves experts working on cleaning and refining the data to ensure it aligns with original publications, then providing feedback to improve the infrastructure. However, this process of data cleaning and manual input is both tedious and time-consuming. In this study, we present expert datasets resulting from systematic data cleaning and input efforts involving 20 geochemists. The high degree of reliability and comprehensiveness achieved in this work provides direct benefits to the research community. To ensure accessibility, the standardized database will be freely available and published as supplements in geochemistry research papers. Additionally, we will continue uploading the standardized data to GEOROC and EarthChem to further improve the data quality of cyberinfrastructures.