Goldschmidt 2022 Abstract
https://doi.org/10.46427/gold2022.12560

OneGeochemistry: Accelerating the Development of Machine Readable Digital Standards for Geoanalytical Data Through International Collaboration

LESLEY ANNE I WYBORN[1], KERSTIN A LEHNERT[2], MARTHE KLÖCKING[3], KIRSTEN ELGER[4], DR. ALEXANDER M PRENT[5], DOMINIK HEZEL[6], LUCY PROFETA[2] AND GEERTJE TER MAAT[7]

[1]Australian National University
[2]Lamont-Doherty Earth Observatory, Columbia University
[3]Georg-August-Universität Göttingen
[4]GFZ German Research Centre for Geosciences
[5]Curtin University
[6]Goethe-Universität Frankfurt am Main
[7]Utrecht University

Presenting Author: lesley.wyborn@anu.edu.au

Since Geochemistry emerged as a new discipline in 1838, vast amounts of geochemical data have been generated by researchers across the globe, across disciplines and from multiple organisations. Today, the volume and speed at which geochemical data are generated is growing exponentially due to new technologies and computerization. Most of this highly valuable data remains hard to reuse and repurpose: much can only be accessed in publications with sparse metadata, or is stored in thousands of globally distributed databases/data systems. There are few agreed standards, vocabularies and conventions to make geochemical data Findable, Accessible, Interoperable and Reusable (FAIR), and readable by both humans and machines. Current human-centric ways of managing, sharing and storing geochemical data no longer scale and are not suitable for new data mining techniques such as AI and ML.

OneGeochemistry is a new international initiative that aims to create global agreement by the geochemical community on standards, vocabularies, formats and conventions for geoanalytical data to enable machine-to-machine data access and processing techniques. The task to generate such standards is enormous. And once developed, a global infrastructure/framework will be needed to ensure that these standards are well documented, governed and accessible online from persistent, sustainable sources.

Fortunately, across the physical and biological sciences many communities have already began addressing the needs for machine-readable (meta)data and have developed standards, vocabularies and conventions that make their data understandable by both humans and machines (e.g., the pure and applied chemistry, crystallography, climate and seismology communities).

But it is not just the natural science community that can help geochemists on their journey. As the global data deluge grew, the new domain-neutral discipline of 'Data Science' emerged that focuses on optimising data for analysis, exploration, understanding and innovation in research: much can be gained from their techniques and experiences.

The OneGeochemistry Initiative is planning to leverage existing processes for developing, publishing, maintaining and revising standards. This presentation will highlight examples in pure and applied chemistry, crystallography, seismology and data science that could be utilised to accelerate the development of the required standards to make global geochemical data FAIR for both humans and machines.