Goldschmidt 2022 Abstract
https://doi.org/10.46427/gold2022.12679

An overview of Hitachi's new data management and analytical modeling platform for geochemistry and automated mineralogy data

JONATHAN P KNAPP[1], DAVID MCKNIGHT[2], ANDREW MUDFORD[2] AND CHAD OSTRANDER[3]

[1]Bruker
[2]Hitachi Vantara
[3]Hitachi High Technologies Canada
Presenting Author: Jonathan.Knapp@bruker.com

The FAIR principles for scientific data require accessible data storage and retrieval systems that are global, accessible, and interoperable. Automated Mineralogy (AM) data has been particularly difficult to provide to end-users due to its complexity and size since it combines imaging and geochemical data to generate mineral maps, consisting of millions of spot analyses and hundreds of high-resolution images per sample. To address this, Hitachi has developed data management and analytical modeling platform for geochemistry that offers a modern purpose-built solution to handle large and complex automated mineralogy datasets with all associated data and metadata on a secure and scalable cloud infrastructure.

This new scientific data infrastructure (SDI) is built around low-latency cloud processing hardware and multi-tiered modern data storage to empower fast and high-volume workflows. The key components are a full project and sample management system, data ingestion from various standard sources, structured storage architecture, pre-built dashboards, analytical tools, and a well-documented and published API interface. Analytical tools include sample management, business user-friendly visualization toolbox, rules-based lithotype classification, image processing, data aggregation engine, and a machine learning framework. The result is a value-adding SDI component for all stakeholders in the scientific data lifecycle to to derive value from advanced analytics and science-driven decision-making.

Hitachi's Lumada Geochemistry Analytics SDI can be an important accelerator for public-private partnerships implementing the FAIR principles. Modern cloud processing and data storage allow for a small per-sample up-front cost, with lower setup and maintenance fees. Costs are incurred based on volume to process and store, not on retrieval frequency or volume, enabling . The system is designed for extensibility and modular readoption, allowing accelerated incorporation of new data models and community resources as add-on “apps” or dashboards or as external computation engines. We challenge the scientific community to adopt clear and applicable FAIR data tools that industry solutions can implement. For example, rich metadata standards around samples, specimens, and measurements make APIs more functional and widely useful. We strongly believe that academia, government, and industry can meet our global data analytics challenges together.