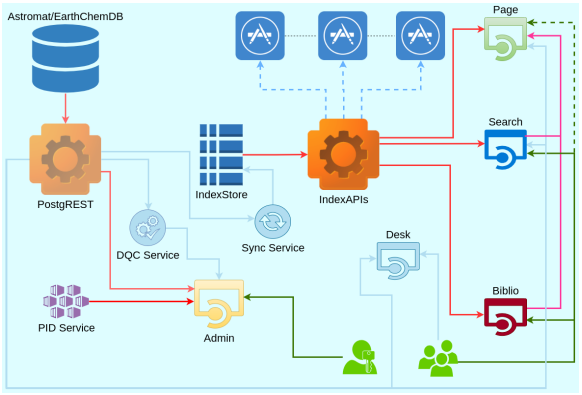
Goldschmidt 2022 Abstract
https://doi.org/10.46427/gold2022.12221

A Shared Cloud-based Architecture for Geochemical Data Systems Applied to Terrestrial and Planetary Materials: The Astromaterials Data System and EarthChem

PENG JI, SEAN CAO, JUAN DAVID FIGUEROA, KERSTIN A LEHNERT AND LUCY PROFETA

Lamont-Doherty Earth Observatory, Columbia University

Presenting Author: pengji@ldeo.columbia.edu



EarthChem and the Astromaterials Data System provide comprehensive data services to research communities that acquire and use laboratory analytical data of terrestrial and planetary materials, respectively, including data publication, data access, data data synthesis, and data preservation. Both systems are developed and operated by the Geoinformatics Research Group (GRG) at the Lamont-Doherty Earth Observatory of Columbia University with the goal to operate a reliable, resilient, scalable, and sustainable data infrastructure in the most efficient and effective manner. Over the past few years, the GRG has created a shared software platform for the two systems that includes service-oriented architecture, public cloud-based infrastructure, and Continuous Integration and Continuous Delivery (CI/CD) pipelines with GitHub Actions (Figure 1 *Service-Oriented Architecture of the Astromat/Earthchem Data System*). Each component of the architecture is intended for specific purposes, including data storage; interfaces for users to search, access, explore, visualize, analyze, and contribute data; software tools for data curators to compile, track, validate, ingest, manage, and annotate data; and machine-actionable interfaces that connect the databases to internal and external software tools.[1] The infrastructure leverages the services provided by AWS, such as S3, Cloudfront, Fargate, etc. to implement the service-oriented architecture illustrated in Fig. 1. Also included is a collection of AWS CloudFormation templates for automating AWS resource management. A series of CI/CD pipelines built with GitHub Actions are used to automate the processes of building, testing, delivering or deploying the applications to AWS infrastructure,resulting in more efficient use of developers' time and effort and greatly improving the reliability of the applications. The shared platform provides the necessary robust reliability and scalability for the data systems, and greatly reduces the need for system maintenance. At the same time, it makes it easier to extend and modify the system to meet the needs of different scientific communities.

Acknowledgments: The work described in this abstract was supported by National Aeronautics & Space [80NSSC19K1102] and National Science Foundation[1948806].

References:

[1] Ji, P., Lehnert, K. New Technology Driving Innovation of Sample Based Chemical Data Systems, American Geophysical Union, Fall Meeting 2019, abstract #IN24A-04.