

Piloting the OSDU platform for inter-state geological survey data management for synthetic 3D mapping initiatives

Project Details

Name of project: Piloting the OSDU platform for inter-state geological survey data management for synthetic 3D mapping initiatives

Project lead and contact details: Gary Motz, Assistant Director for Information Services

Indiana Geological and Water Survey, Indiana University

garymotz@indiana.edu, 812-856-3500

Project partners and contact details: Denise Hills, Director of Energy Investigations, Geological Survey of

Alabama, dhills@gsa.state.al.us; Jessica Czajkowski, Assistant Director of Science and Research,

Washington Geological Survey, jessica.czajkowski@dnr.wa.gov; Doug Curl, Head of Geoscience

Information Management, Kentucky Geological Survey, doug@uky.edu; Drew Andrews, Head of

Geologic Mapping, Kentucky Geological Survey, wjandr00@email.uky.edu; Marcella McIntyre,

Energy Investigations Scientist, Geological Survey of Alabama, mmcintyre@gsa.state.al.us

Proposed start and end date: July 1, 2021 - January 31, 2022 (7 months)

*Budget Requested: \$9,750 traditional funding, \$5,000 AWS cloud services, and data mentorship

Project Outline

Project description: “*Modernization of Earth science workflows using community-recommended best practices — the use of open-source software and cloud computing are encouraged.*” This ESIP Lab funding priority, in particular, stuck out to us as quite relevant to how we want to do business as an organizational priority and pivot our institutional data management practice toward a data-centric infrastructure that leverages both open-source software and cloud computing. The Open Subsurface Data Universe (OSDU) is a platform built and maintained with open, international standards and schemas by The Open Group (The Open Group, 2020). In January 2021, the Indiana Geological and Water Survey (IGWS) affirmed its commitment to FAIR data practices, standards-based processes, and data-informed decision support infrastructure in its revised strategic plan for the organization as a whole and to be implemented over the next five years (when finalized, will be publicly available at <https://igws.indiana.edu/more-information>). To accomplish these strategic objectives, an implementation plan that calls for a targeted assessment of individual products and platforms that may facilitate these goals for the IGWS data re-architecture is now underway, including the effort contained within this proposal and extended by collaboration with peers at other state geological surveys.

Through this proposed ESIP Lab project, we intend to investigate two primary questions: How can we apply standards-based data management ecosystems to build data-proximate collaborative networks within the OSDU to facilitate enhanced interoperability of subsurface data to promote greater reconciliation of data between states for geologic mapping and derivative data products? Is there a substantive difference between the results of collaborative, seamless mapping using the nascent Geological Mapping Schema standard (GeMS, USGS NCGMP, 2020), or do OSDU standards promote a greater degree of data interoperability between adjacent states?

Project objectives, significance, and impact: The proposed project has three objectives that are to be achieved by this project:

- initializing an instance of the Open Subsurface Data Universe cloud-native platform on Amazon Web Services (AWS) for subsurface data management and sharing this environment in a Platform-as-a-Service model for replication by other state geological surveys
- building a comparison model of geological mapping efforts across state/institutional boundaries to assess the degree of interoperability between two different standard sets
- determining the degree to which the USGS Geological Map Schema aligns with OSDU-facilitated standards and set a baseline for future standards development

To date, the OSDU is a relatively novel data management environment that has seen broad adoption in the energy and exploration sector of commercial industry and is slowly making its way into the everyday vocabulary of academic and research institutions, including state geological surveys. Substantial investment in the OSDU as a data management environment *specifically* for multidimensional relationships between numerous subsurface data sources in an earth science context by multinational corporations. This open, standards-based initiative aimed to produce a purpose-built, cloud-native platform for earth science data management instead of co-opting a data management platform designed for other communities of practice and then heavily customized to work for earth science purposes. This project proposes, to the best of our knowledge, the first foray into a discovery of the utility of the Open Subsurface Data Universe by any geological survey for geological mapping and research purposes.

Description of key project steps and timeline:

1. July 2021 – clone OSDU on AWS from GitLabs into ESIP Lab project repo and provision OSDU
2. Aug 2021 – customize OSDU deployment and distribute to project partners
3. Sept – Nov 2021 – input data, test database interactions, facilitate schema alignment between state geological survey local implementations, GeMS, and OSDU schemas
4. Oct – Dec 2021 – Compare notes with participating geological surveys
5. Jan 2022 – compile and present findings at ESIP Winter Meeting

Description of additional funding currently supporting this work:

The IGWS has already committed funds from the appropriation that the State of Indiana allocates to the Indiana Geological and Water Survey to support this initiative and submitted external funding requests to the USGS STATEMAP program to enhance enterprise data management for geological mapping.

Outreach

What groups/audiences will be engaged in the project?

We aim to engage the four state geological survey partners named in this proposal, the American Association of State Geologists, the ESIP Physical Samples Cluster, the USGS Community for Data Integration, the USGS National Geological and Geophysical Data Preservation Program, the USGS Data Management Working Group, the USGS National Geologic Map Database committee, and share detailed presentations of our findings with these communities of practice. Each of the state geological survey participants in this project will contribute data to a cloned testbed of the OSDU architecture platform on an AWS instance spun up by the IGWS. After each platform clone is initialized, IGWS, GSA, KGS, and WGS personnel will evaluate the suitability of their individual database structures and the degree to which data can be imported within the OSDU framework.

How will you judge the project's impact?

Initial interest in testing the efficacy of the OSDU environment has been expressed by organizations beyond the four state geological surveys that have joined this project. The impact of this project will be readily assessed by determining whether and how many additional state geological surveys or the USGS may follow suit and leverage the documentation and experiences shared by this pilot.

How will you share the knowledge generated by the project?

The project participants will share knowledge generated by this project by contributing a well-commented GitHub repository with deployment and configuration instructions committed to the ESIP Lab GitHub repository established for this project. Additionally, the project participants will present the findings of this project at both the 2022 Winter ESIP meeting and a meeting of the newly-formed ESIP Physical Samples cluster.

Project Partners

This project aims to undertake a critical evaluation of the suitability of the OSDU as a standards-based data management platform for subsurface data. The Open Group has initiated a broad schema of standards developed for and by energy and exploration companies. Still, the interoperability of data maintenance natively in multidimensional relationships engenders a novel approach that may be suitable for broad adoption by the community of geological surveys across the country. These data standards in a cloud-

native cyberinfrastructure are designed to facilitate enhanced silo reduction, automated metadata parsing and indexing, and structured relationships that enable AI-powered analytics across highly diverse and heterogeneous datasets.

A major synthesis project coordinated between USGS National Cooperative Geologic Mapping Program and the state geological surveys via AASG (the American Association of State Geologists) is to facilitate a seamless 3D geological map of the nation. To date, implementation of the GeMS standard in the construction of geodatabases for map publications has been rolled out in three phases of compliance or adherence with a prescriptive format in the metadata schema. Individual state geological surveys are, in general, beginning to implement the third degree of compliance implementation within state boundaries: few interstate or regional compilations of geological map data and underlying source data structures yet exist. This project will lay the groundwork for a standards-based framework and data architecture to promote enhanced openness in the most fundamental work of any geological survey.

Budget

\$700 – Presentation of project findings and outcomes at Winter 2022 ESIP meeting

\$3000 – Initializing containerized OSDU architecture for replication across testbeds in each participating state geological survey

\$3025 – Build comparison model of state line map congruence across Indiana/Kentucky border with GeMS-aligned geodatabase in contrast with OSDU implemented data architecture

\$3025 – Assessment of interoperability of Geologic Map Schema (GeMS) elements with OSDU metadata schema and elements

References

Amazon Web Services, 2021, OSDU Data Platform on AWS website, at URL:

<https://aws.amazon.com/energy/osdu-data-platform/?nc=sn&loc=5>, date accessed, January 25, 2021.

Indiana Geological Survey (IGS), 2016, Petroleum Database Management System: URL at

<https://igws.indiana.edu/pdms/Map/>, date accessed, January 15, 2021.

McLaughlin, P. I., Baker, M., and Johnson, M. R., 2017, Bedrock Core Information System: Indiana Geological and Water Survey Digital Information Series 17, URL at <https://igws.indiana.edu/IGSMap/BedrockCoreInfo>, date accessed, January 15, 2021.

The Open Group, 2020, Open Subsurface Data Universe software platform deployment website, at URL:

<https://community.opengroup.org/osdu/platform/deployment-and-operations/infra-azure-provisioning>, date accessed, December 15, 2020.

U.S. Geological Survey National Cooperative Geologic Mapping Program, 2020, GeMS (Geologic Map Schema)—A standard format for the digital publication of geologic maps: U.S. Geological Survey Techniques and Methods, book 11, chap. B10, 74 p., <https://doi.org/10.3133/tm11B10>.