CEOS Earth Analytics Interoperability Lab

Working Group on Information Systems and Services and Systems Engineering Office

20 April 2020
Robert Woodcock, CSIRO, WGISS Chair
Brian Killough, NASA, SEO Director

Situation

A significant number of CEOS activities are now engaged in the CEOS ARD and FDA strategies and in Integrated Earth observation data analysis (COAST, SDG, WGClimate, LSI-VC, WGDisasters, GEO Aquawatch and GEOGLAM). There is strong collaboration between groups with both WGISS and SEO being sought for technical advice and coordination on issues related to interoperability of all kinds including data standards, formats, analytics and discovery services.

To date coordination has taken the form of project linkages, development of common terminology (e.g. CEOS Interoperability Terminology Report), and joint development and use of best practice approaches (e.g. The LSI-VC CARD4L specification and process is being used as a template for the for CARD4Water). This has worked well for more isolated outputs like data standards.

Current project activities are seeking to improve data discovery and analytics interoperability with often subtle impacts on CEOS services be they provided by WGISS, SEO or CEOS agencies. The CEOS community has reached a point where ARD and FDA activities need to become demonstrable and testable in a live service ecosystem in order to validate interoperability throughout the EO value chain from Discovery through to Analytics outcome.

Opportunity

Validating interoperability between multiple CEOS organisations and working groups is a complex coordination effort. Typically, standards groups use interoperability Testbed activities which operate outside of organisational constraints to support the testing of new ideas from multiple groups until a clear approach is validated.

There is an opportunity for WGISS and the SEO to jointly provision a CEOS Earth Analytics Interoperability Lab to provide a platform for CEOS projects to test interoperability in a live EO ecosystem. The Lab would be hosted by SEO and jointly operated by WGISS and SEO to provide:

- Examples of FDA components in active use including new Discoverability and access approaches (e.g. connection to USGS and ESA data in the Cloud)
- A shared ARD storage and access capability for candidate ARD outcomes and comparative analysis by CEOS teams to support ARD validation
- Connection to existing and emergent CEOS and Agency services for Discoverability and Access
- Collaboration on Analytics tools for integrated analysis, including sample data, using Jupyter Notebooks

Without such a shared platform CEOS work program activities will build their own baseline systems and which operate in isolation (as they are most likely provided by individual agencies). The Lab would reduce this requirement allowing greater participation and evaluation of interoperability outcomes in an open environment, until such time as the interoperability requirements are understood and agencies can implement them.

Individual agencies often have the capability to provide such services. This proposal does not seek to circumvent these efforts, as ultimately agencies will be providing all services. The Lab does seek to provide a stepping-stone to support the coordinated and collaborative development of interoperable services in order to develop best practices that agencies can more readily adopt.

Proposal

WGISS and the SEO propose building the platform out from the existing CEOS Open Data Cube infrastructure. The initial proposal is to add the CSIRO EASI Hub environment to the CEOS Open Data Cube and create the first version of the CEOS Interoperability Lab. The CSIRO EASI Hub includes facilities for Jypyter notebooks, scalable compute infrastructure for integrated analysis and data pipelines which can connect to new and existing CEOS data discovery and access services. The CEOS interoperability Lab would be hosted on AWS, consistent with the existing CEOS SEO Open Data Cube and support modest amounts of data and analytics necessary for interoperability experiments in the CEOS projects.
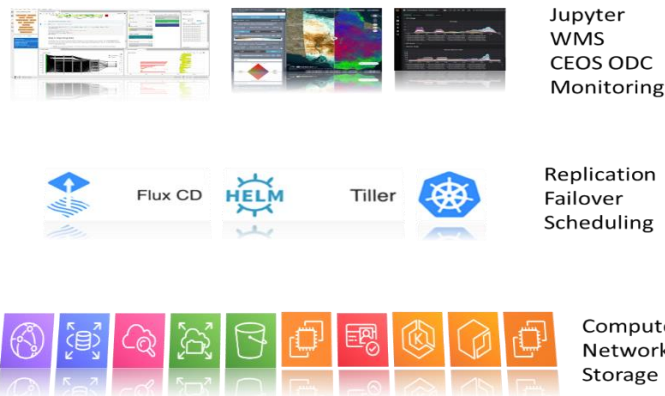
Discussions with other WG and VC activities and CEOS projects suggest this would provide the necessary base feature set to support their initial development. It is recognised that moving to such an environment will require some capability development. WGCapD has already been exploring the use of Jupyter Notebooks and is also seeking to develop new course work and provide hands on training in more the emergent areas of ARD and FDA. Initial discussions suggest WGCapD may be able to use the same facility for the development of such courses and assist in developing necessary material for the Lab.

As the ecosystem evolves additional capabilities will likely be added by other CEOS agencies as part of interoperability experiments (e.g. ESA has both data cube and jupyter notebook capability). It is precisely this evolution which will ultimately achieve the high degree of interoperability that is being sought to achieve CEOS objectives.

Feedback and contributions to this proposal are welcome.

Robert Woodcock - WGISS Chair

Brian Killough – SEO Director