EOSDIS cloud evolution and STAC

Directive

The ability to portray our cloud-based data holdings as SpatioTemporal Asset Catalogs (STAC) is an essential component of the ESDIS cloud evolution strategy.

Solution

The Common Metadata Repository will provide a STAC API that will allow discovery of all ESDIS cloud-based holdings, subject to established access controls and suitability[1].

All STAC-suitable data must have corresponding metadata in the CMR that allows CMR to generate a STAC catalog from that data.

Conformance

The CMR STAC API shall be conformant with version 1.0.0 of the STAC API specification and version 1.0.0 of the STAC Catalog.

The adherence to those specifications and any mandated extensions (and future mandated versions thereof) should be demonstrable via testing.

CMR STAC will be directed to be compliant with necessary STAC extension specifications at ESDIS' discretion.

Evolution

CMR STAC will be directed to be compliant with major updates of the STAC Catalog, STAC API and necessary extension specifications at ESDIS' discretion.

Use cases

1. As a science user in a notebook environment, I would like to use STAC libraries to locate STAC catalogs representing my data of interest in the cloud. I would use those catalogs to interact with standard geospatial analysis libraries such as GeoPandas, intake, XArray and DASK to produce efficient, cost-effective, reproducible analysis. This notebook environment may be local or cloud-based.
2. As a science user, I would like to discover data of interest using STAC tooling such as STACBrowser.

Improvements

ESDIS has a commitment to providing an efficient, intuitive user experience when leveraging our holdings. Consequently, we have identified several areas with respect to STAC that could be improved in this regard.

While these are not showstoppers for cloud evolution with respect to STAC, we believe they should be addressed in the medium term.

Authentication and access control

In order to provide STAC catalogs for our entire cloud-based inventory, we need to provide restricted access to non-public data. CMR STAC needs to allow a user to provide credentials to access such data, if allowed. CMR STAC must provide the same level of authorization that the native CMR search API provides. [https://bugs.earthdata.nasa.gov/browse/CMR-7104] Any provision of access control should be in line with STAC recommendations should they exist now or in the future.

Bringing compute to the data - direct S3 access

In order to access our holdings, a user needs to authenticate. For in-region direct access, that involves the use of short term AWS credentials rather than EDL credentials. But those AWS credentials are dispensed through the use of EDL credentials.

Our current user experience for facilitating the above is geared towards leveraging the UMM.

The direct S3 access workflow involves the following,

1. A user/application discovers a collection and granules of interest
2. The granule metadata describes S3 URLs for direct access
3. The collection metadata describes the S3 credentials endpoint for that data.
   - S3 credentials endpoint
   - S3 credentials documentation - ie. supplying your EDL credentials to get temporary AWS credentials
4. The user/application obtains AWS credentials using their EDL credentials
5. The user/application uses those AWS credentials to access the data using the AWS S3 API.

There is, currently, no support for obtaining AWS credentials for the S3 API in the STAC Catalog/API specification or the CMR STAC API implementation.

https://github.com/stac-extensions/storage and https://github.com/stac-extensions/alternate-assets could be improved upon to add credential information.

The S3 credentials improvements can be mitigated by outreach and documentation in the short term.

1 Some ESDIS data is not suitable for STAC. Those data without spatial and temporal extents, for example