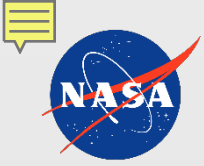


CMR-STAC

Overview

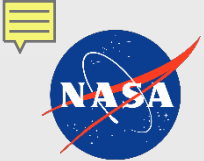
September 2020

Valerie Dixon
NASA ESDIS
valerie.dixon@nasa.gov



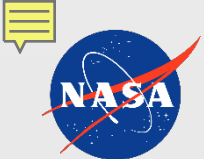
History and Context

- In 2018, OGC's TestBed 14 highlights utility of WFS 3.0
 - moves away from XML
 - Implements a simple RESTful interface, primarily using JSON and HTML, with OpenAPI support
 - Making it much faster and easier to programmatically create and read requests
 - With WFS's success, WCS and WMS are already following suit
- Shortly after, Radiant Earth's Spatio-Temporal Asset Catalog (STAC) came along to exploit this new query functionality
 - Search & discovery-centric JSON-defined metadata
 - Simple and extensible
 - Dynamic APIs to search arbitrary fields
 - Static, crawl-able catalogs



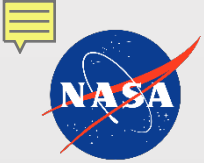
STAC & CMR

- ESDIS paired with Element84 to develop a STAC API Proxy for CMR
 - Translates requests/responses between STAC endpoints and CMR
 - So CMR providers don't have to change their metadata
 - CMR-STAC implements ALL of core STAC, with additional extensions for:
 - Content: extra metadata fields for particular types of data, like bands for EO data
 - API: for additional endpoints or arguments to enable one to search for/return particular fields
 - CMR STAC [Documentation](#) and [GitHub](#) links
- CMR STAC as an *API Proxy*...
 - Supports crawling through the data in a browser via primary [CMR-STAC endpoint](#)
 - Supports users making direct HTTP requests to provider endpoints
 - Endpoints for GET and POST requests designed to be used by third-party tools like [Sat-Search](#)
- ...is not a *static STAC catalog*
 - Static STAC catalogs enable more in-depth browsing, but extensions are closing the gap



STAC Limitations & Uses

- The STAC metadata model is NOT a complete metadata standard
 - Too Simple
 - Doesn't cover all kinds of datasets or metadata fields
 - Still Evolving
- But it IS a useful spatiotemporal query tool
 - That embeds nicely with Jupyter Notebooks and Python
 - That can query for specific scenes, granules, and potentially even bands
 - That has a growing and enthusiastic user base
 - Summer 2020 ESIP session all about STAC and Sentinel-2, COGs, Pangeo
 - And a growing number of third-party applications and libraries to support it
 - [Intake-STAC](#), [Sat-STAC](#), [Pystac](#), [QGIS-STAC-browser](#), etc.



Recent Innovations in CMR-STAC

- CMR-STAC endpoints create a STAC catalog for HLS data for use in Pangeo
 - In Production now, resolving the following issues:
 - Collections:
 - Pagination in collections
 - Temporal extent incorrect for ongoing collections
 - Reorder spatial extents
 - Collection Validation
 - Incorrect media types for metadata
 - Proper media types for links
 - Remove double slash from hierarchical links
 - Items
 - Items and Search endpoints not in ItemCollections
 - Items missing required fields 'bbox' and 'stac_version'
 - Move metadata from 'links' to 'assets'
 - Links should be a list, not dictionary
 - Reformat Datetime field to conform to STAC Spec
 - Collection ID and collection link missing from Items
 - Browse image media types
 - Remove unnecessary '?' characters from URLs
 - Item pagination should use 'rel=next'
 - And adding the following new functionality:
 - Upgraded to STAC v1.0.0-beta.1
 - More thorough response validation
 - More robust error handling
 - Improved Search support
 - Restructured endpoints



Looking ahead...



- EO and SAR extensions for CMR-STAC response (Sentinel, MAAP)
 - May result in updates to UMM to add extensions for things like SAR
 - Could also be done using additional attributes, granule level tags, or by injecting at the STAC translation layer
- Analysis-in-Place
 - Could STAC play a role? Need to analyze end to end UX to find out.
- Link between STAC Features and Coverages
 - Upgrade CMR STAC to keep current with latest OGC-Feature version
 - WFS is for features and we tend to focus on coverages. Each STAC record is a geoJSON (feature) and references underlying data which can be used as coverages.
 - Basic write-up of relationship between STAC, Features, Coverages
- Document how to handle data deletions and replacements
- Develop further STAC extension for provenance workflow
 - Could use STAC file moves w/ data to get provenance.



Looking ahead, continued...



- Further Pangeo integration/support
 - COGs, JupyterHub, Kubernetes, Python, ZARR, etc.
- OPeNDAP (DMR++) top-level STAC records/virtual data cube for NetCDF files
 - Aggregate a dataset over time w/ OPeNDAP (DMR++) and surface via STAC
 - Previously really, really hard with NCML; DMR++ is an order of magnitude better
 - Still have NetCDF files natively
- CMR-STAC applicability in disaster response data cubes
- Add full STAC-Browser capability
 - Maybe through making CMR's Virtual Directory API into/analogous to a static STAC catalog?
- Harmony tie-in
 - Staged harmony transformation results should be/have a STAC catalog, so tooling can work directly with it
- Investigate relationship w/ Landsat and AWS public datasets



Special Thanks

- Element84

Travis Vance tvance@element84.com

Drew Daniel drew@element84.com

Daniel Pilone daniel.j.pilone@nasa.gov

Matthew Hanson mhanson@element84.com

Dylan Clark dylan@carbonfive.com

- EED

Joseph Rincione joseph.rincione@nasa.gov

Doug Newman douglas.j.newman@nasa.gov