Intra-agency 'Data Commons' (a node in the 3DEP Data Mesh Ecosystem)

Deployment

Terraform script

- Allocates cloud compute, storage, security groups.

Domain Initialization

GSA Customization 1:

OAuth2/IAM/AD/ID.me

Self-reported ID 🖪

(Stub for PoC)

- Via Dockerfile RUN commands
- Initial setup/ingest of static resources
- Optionally regester this Data Commons instance with catalog server
- Or manually add other external Data Commons for data discovery

STAC API IIIII

Server Components (Docker composed together)

- Real-time, dynamic querying, filtering, and retrieval of data
- provides RESTful interface to intra-agency STAC catalog
- implemented via Python/FastAPI

STAC Browser

- implemented via

pgSTAC/PostGIS

- Browser-based data discovery tool
- Search for data products both internal and external to local STAC catalog

Intra-agency STAC Catalog

 Kick off new processing workflows with search results

JupyterHub 🏝

- Can receive discovered search results as input with blank workspace
- Example Jupyter notebooks provided
- 3DEP libraries preinstalled (Entwine, PDAL, GDAL)

GSA PoC Customization of STAC API: RegisterDataProduct()

Baseline API endpoint definition:

- STAC API Extension Endpoint Transaction' - Supports HTTP methods POST, PUT, PATCH, DELETE - Input body: 'STAC Item' GeoJSON Feature augmented with 'Foreign Members'

GSA Customization 2: Provenance

- Enforce existence of 'Link' object of relation type='derived_from'
- could be upstream data service endpoint URL, filename, REST API call, GUI-based map service, etc

GSA Customization 3: Data Product Info

- Jupyter notebook filename/URL (register new notebook now?)
- Source code has to be referenceable
- Plain language description
- return value type of data product (Stub for PoC)

Legend

Four Pillars of Data Mesh:

- № Domain-Oriented Decentralized Data Ownership
- IIII Data as a Product
- Self-service Data Infrastructure as a Platform
- 🖪 Federated Computational Governance

Hover over/click boxes for links to definitions.