

Data Mesh Component 1: Intra-domain STAC Catalog & Jupyter server ('Walled Garden')

Deployment

Terraform script ☁️

- Allocates walled garden compute, storage, security groups.
- 📅 Phase 3

Domain Initialization ☁️

- Via Dockerfile RUN commands
- Initial setup/ingest of static resources
- Register this domain with the central catalog of catalogs
- 📅 Phase 2

Server Components (Docker composed together)

Domain STAC Catalog 🗄️

- implemented via Pystac/static file or pgSTAC/Postgres 🗄️
- 📅 Phase 1

STAC API 🗄️

- provides RESTful interface to domain STAC catalog
- implemented via Python/FastAPI Docker container
- 📅 Phase 1

STAC Browser GUI 🗄️

- GUI front-end querying local STAC REST API
- 📅 Phase 2

JupyterHub ☁️

- Run OpenTopo Jupyter notebooks
- long-term: Kubernetes or Databricks-backed
- 📅 Phase 3

Extension to STAC API Transaction Endpoint: Register()

Required: Self-reported ID 🗄️

OAuth2/IAM/AD/ID.me (Stub for PoC)

Required: Provenance 🗄️

- Upstream Data Service Endpoint URL
- filename
- REST API call
- Some GUI-based map service

Required: Data Product Info 🗄️

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

Legend

Four Pillars of Data Mesh:

- 🗄️ Domain-Oriented Decentralized Data Ownership
- 🗄️ Data as a Product
- ☁️ Self-service Data Infrastructure as a Platform
- 🗄️ Federated Computational Governance

GSA PoC Feature Milestone Phases:

1. 📅 February 2024
2. 📅 March 2024
3. 📅 April 2024