EOSDIS
NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Connecting Data with Data Usage: A Graph Approach

Winter ESIP, 10:15am, Thursday, January 9th 2020

Doug Newman
NASA EED-2 Data Use Architect
douglas.j.newman@nasa.gov

Dr. Christopher Lynnes
NASA EOSDIS System Architect
christopher.s.lynnes@nasa.gov

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.

INTRODUCTION

The rationale

‘Connect together the main elements of Earth Observation knowledge AND context in a way that is: machine-readable, human-usable and curatable’

Chris Lynnes, WGISS-47

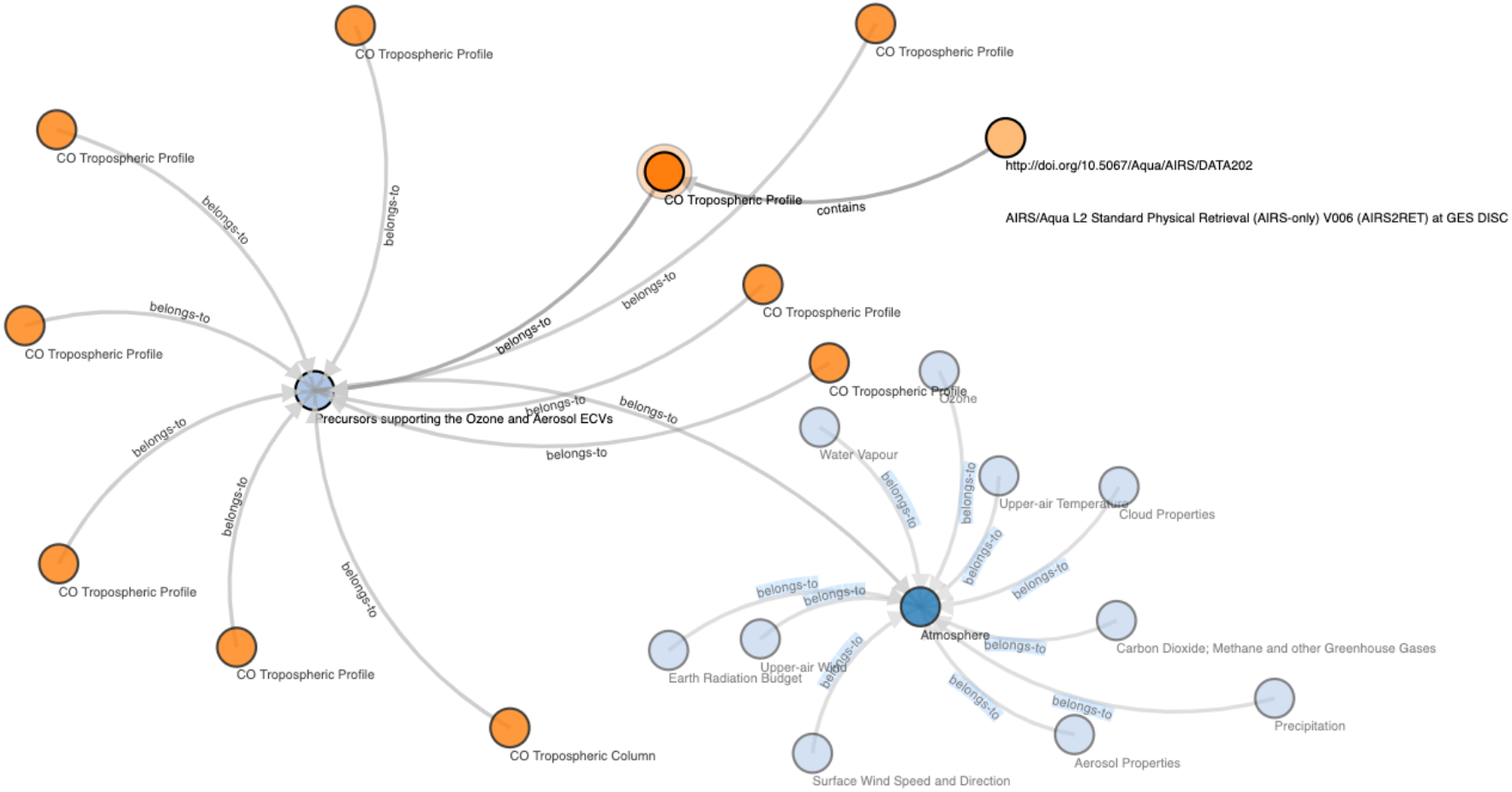
Or...

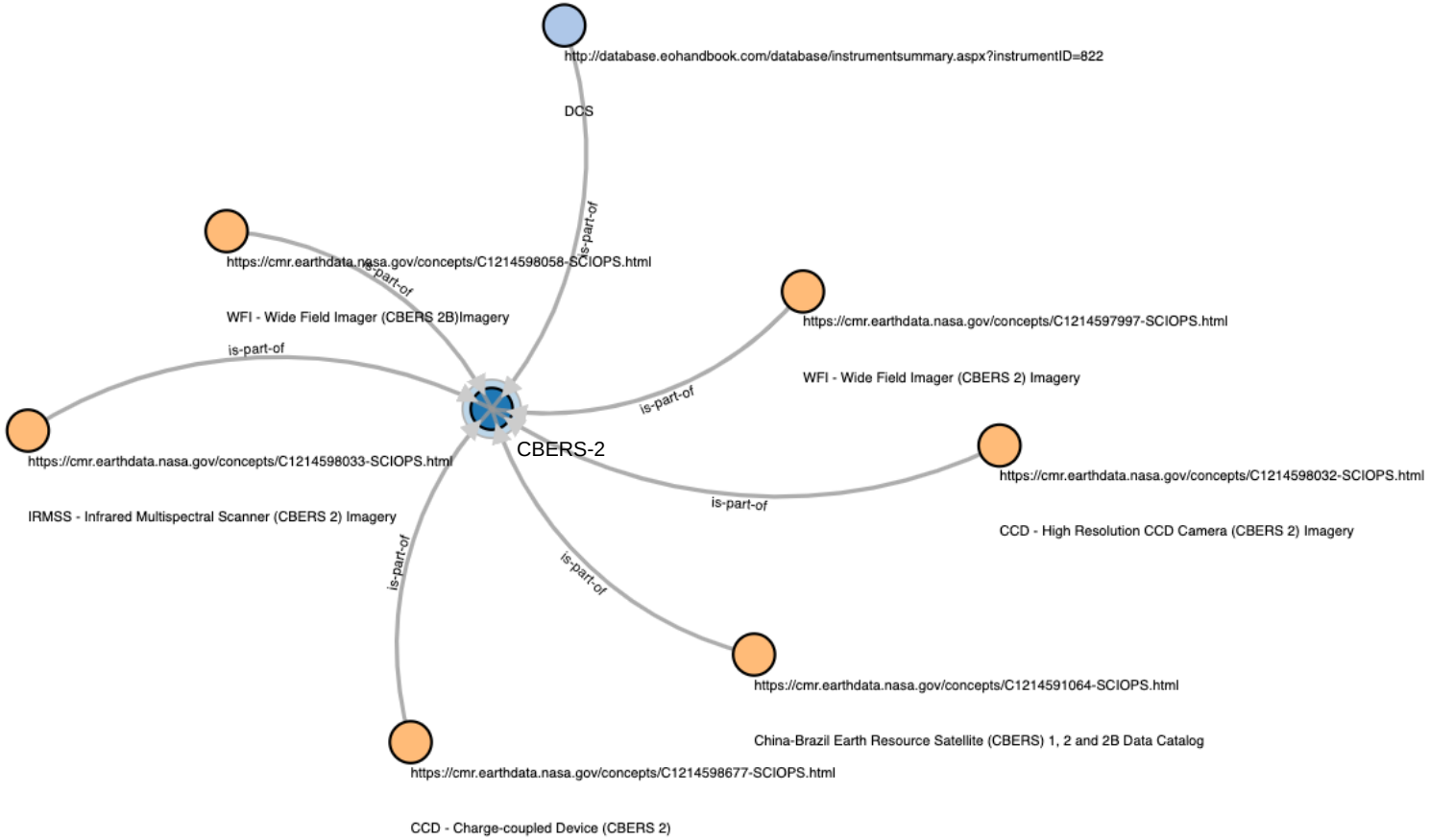
‘Scientists are swimming in a sea of datasets and want to know which ones to use.’

Chris Lynnes, numerous

Quick overview

- **Elements of knowledge:** Missions, instruments, datasets, measurements, articles
- **Context of knowledge:** this instrument belongs to this mission, this dataset contains this measurement, this article cites this dataset
- All this can be modeled efficiently and intuitively using graph concepts


CO Tropospheric Profile
CO Tropospheric Profile
CO Tropospheric Profile
CO Tropospheric Profile
http://doi.org/10.5067/Aqua/AIRS/DATA202
CO Tropospheric Profile
contains
AIRS/Aqua L2 Standard Physical Retrieval (AIRS-only) V006 (AIRS2RET) at GES DISC
belongs-to
CO Tropospheric Profile
CO Tropospheric Profile
CO Tropospheric Profile
Precursors supporting the Ozone and Aerosol ECVs
CO Tropospheric Profile
Ozone
Water Vapour
Upper-air Temperature
Cloud Properties
CO Tropospheric Profile
CO Tropospheric Profile
CO Tropospheric Column
Atmosphere
Earth Radiation Budget
Upper-air Wind
Carbon Dioxide; Methane and other Greenhouse Gases
Precipitation
Aerosol Properties
Surface Wind Speed and Direction


http://database.eohandbook.com/database/instrumentsummary.aspx?instrumentID=822
DCS
is-part-of
https://cmr.earthdata.nasa.gov/concepts/C1214598058-SCIOPS.html
is-part-of
WFI - Wide Field Imager (CBERS 2B)Imagery
is-part-of
https://cmr.earthdata.nasa.gov/concepts/C1214597997-SCIOPS.html
WFI - Wide Field Imager (CBERS 2) Imagery
is-part-of
https://cmr.earthdata.nasa.gov/concepts/C1214598033-SCIOPS.html
IRMSS - Infrared Multispectral Scanner (CBERS 2) Imagery
CBERS-2
is-part-of
https://cmr.earthdata.nasa.gov/concepts/C1214598032-SCIOPS.html
CCD - High Resolution CCD Camera (CBERS 2) Imagery
is-part-of
is-part-of
https://cmr.earthdata.nasa.gov/concepts/C1214591064-SCIOPS.html
China-Brazil Earth Resource Satellite (CBERS) 1, 2 and 2B Data Catalog
is-part-of
https://cmr.earthdata.nasa.gov/concepts/C1214598677-SCIOPS.html
CCD - Charge-coupled Device (CBERS 2)

NASA Earthdata

Overview

- Effective technologies
- Sources of knowledge
  - NASA EOSDIS Common Metadata Repository
  - World meteorological organization
  - The CEOS database

Graph Technologies

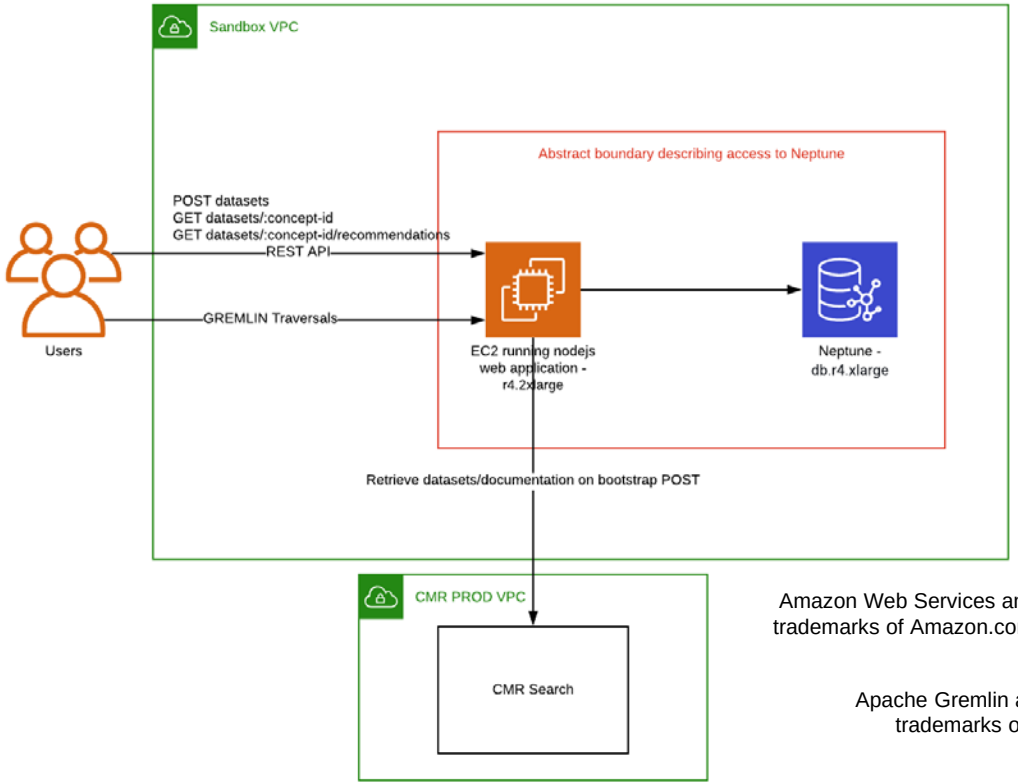
Implementations

- AWS Neptune
- Neo4J

APIs

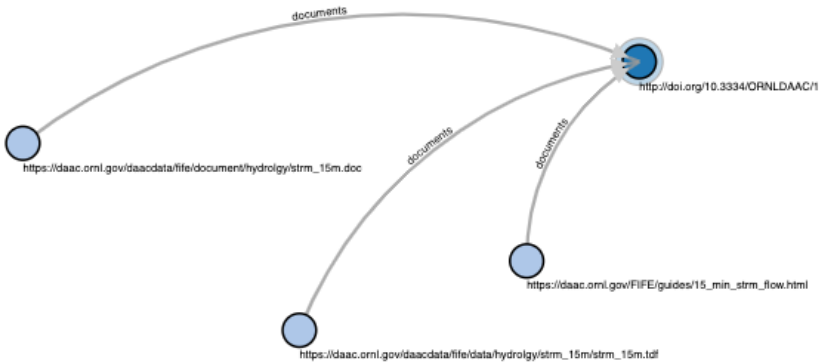
- SPARQL
- Gremlin
- Cypher - now GQL

AWS Neptune + Gremlin

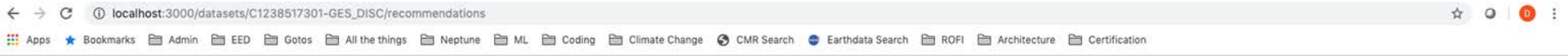


Within NASA Earthdata

The common metadata repository contains dataset metadata records and references to their respective documentation. We used this inventory to create a graph of datasets and documentation vertices and ‘documents’ edges, originally to serve as a stepping stone towards scientific literature.

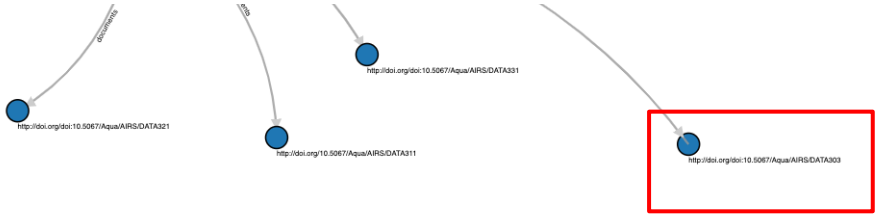


Side trip to recommendations



Similar datasets to 'AIRS/Aqua L3 Monthly Standard Physical Retrieval (AIRS-only) 1 degree x 1 degree V006 (AIRS3STM) at GES DISC'

- AIRS/Aqua L3 8-day Standard Physical Retrieval (AIRS-only) 1 degree X 1 degree V006 (AIRS3ST8) at GES DISC
- AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS+AMSU+HSB) 1 degree x 1 degree V006 (AIRH3STD) at GES DISC
- AIRS/Aqua L3 8-day Standard Physical Retrieval (AIRS+AMSU+HSB) 1 degree x 1 degree V006 (AIRH3ST8) at GES DISC
- AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS+AMSU) 1 degree x 1 degree V006 (AIRX3STD) at GES DISC
- AIRS/Aqua L3 Monthly Standard Physical Retrieval (AIRS+AMSU+HSB) 1 degree x 1 degree V006 (AIRH3STM) at GES DISC
- AIRS/Aqua L3 Daily Standard Physical Retrieval (AIRS-only) 1 degree x 1 degree V006 (AIRS3STD) at GES DISC
- AIRS/Aqua L3 5-day Quantization in Physical Units (AIRS+AMSU) 5 degrees x 5 degrees V006 (AIRX3QP5) at GES DISC
- AIRS/Aqua L3 8-day Standard Physical Retrieval (AIRS+AMSU) 1 degree x 1 degree V006 (AIRX3ST8) at GES DISC
- AIRS/Aqua L3 Monthly Standard Physical Retrieval (AIRS+AMSU) 1 degree x 1 degree V006 (AIRX3STM) at GES DISC



AIRS3STM

```
g.V().hasLabel('dataset')
.has('conceptid',conceptId)
.inE('documents')
.outV()
.hasLabel('documentation')
.outE('documents')
.inV()
.hasLabel('dataset')
```

Essential Climate Variables (1)

- A good starting point for discovery: ‘I want *total ozone* data’
- Total Ozone is one of 496 well-defined products, spread across 30 variables in 3 domains

Essential Climate Variables (2)

425 datasets have been mapped to those products in our graph implementation

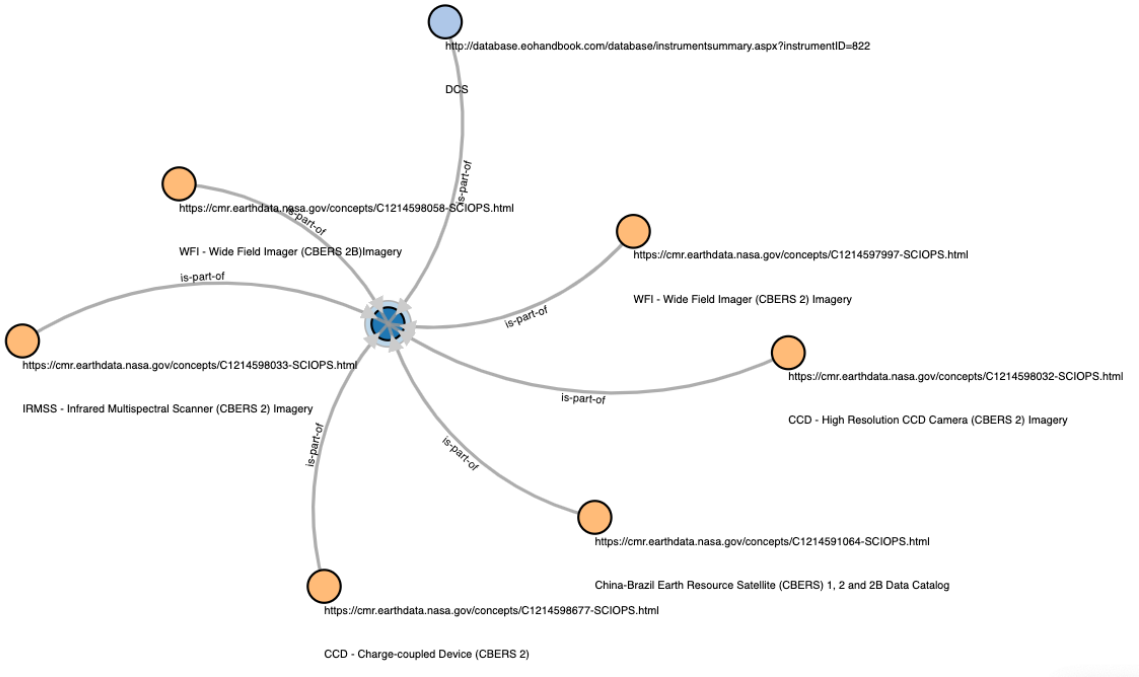
Essential Climate Variables (2)

Missions and instruments

The CEOS database contains extensive information about missions and instruments and their linkage.

CMR contains linkage information between datasets and mission/instruments.

Missions and instruments



Keywords

The hierarchy of science keywords and their associations with datasets

- Science keywords - GCMD
- Datasets - CMR

Roadmap

- Graph implementation in EED by Q2 2020
  - CMR concept associations
- ECV mappings for GES_DISC by Q2 2020
- ECV graph implementation by Q3 2020

CONCLUSIONS

NASA: If you build it...

- 10 minutes looking at a visualization of two CMR concepts gave me the inspiration for the recommendation engine. Imagine what an expert could do?
- Providing a read-only traversal API over a knowledge base containing more concepts could open the floodgates for novel discovery techniques

NEXT STEPS

Link all the things

- With help from each other we would like to scale these data connections to facilitate data discovery
- Graph traversal APIs will enable novel discovery techniques

QUESTIONS?

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.

Raytheon

in partnership with

ati

Bellfore Consulting Group

Element 84

Genex Systems

The HDF Group

Innovim

Maryland Hawk Corporation

OPeNDAP

Red Canyon Engineering & Software

Stinger Ghaffarian Technologies

Valador