Semantic Web Applications

at the Goddard Earth Sciences Data and Information Services Center (GES DISC)

Christopher Lynnes
Greg Leptoukh

The Killer App?

- User: “Run a comparison of satellite-based aerosol optical depth measurements for the 2008 Popocatepetl volcanic eruption”
- Application
  - Go find aerosol optical depth measurement sources, aka AOD, aka optical depth
  - Distinguish between total column AOD (yes) and troposphere AOD (no)
  - Distinguish between ground-based and satellite-based
  - Find comparable datasets
  - Find and interpret the fields that specify quality

Baby Steps

- Mirador search tool
  - Ontology used to structure hierarchical navigation
- Multi-Sensor Data Synergy Advisor
  - Ontology provides controlled vocabulary
- **Data Quality Screening Service**
- Aerostat
  - Ontology for bias description?
  - Ontology for results tagging?

Satellite Data Quality: Why So Difficult?

Satellite-borne collection breeds a tendency to keep low-quality data around

- Satellites are *very expensive* to launch and operate.
- It takes years to understand the data characteristics.
- Data are used for many different purposes.

AIRS – Atmospheric Infrared Sounder

- atmospheric temperature
- atmospheric moisture
- trace gas composition

Current user scenarios…

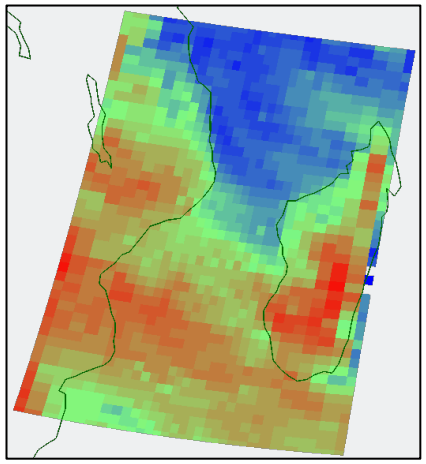
- Nominal scenario
  - Search for and download data
  - Locate documentation on handling quality
  - Read & understand documentation on quality
  - Write custom routine to filter out bad pixels

Repeat for each user

- Equally likely scenario (especially in certain user communities)
  - Search for and download data
  - Assume that quality has a negligible effect

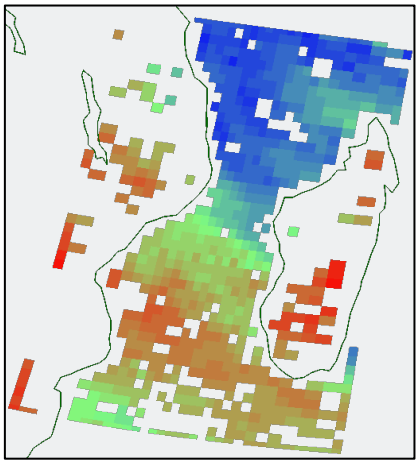
DQSS segregates bad-quality pixels into a separate array

Original data array (Total column precipitable water)

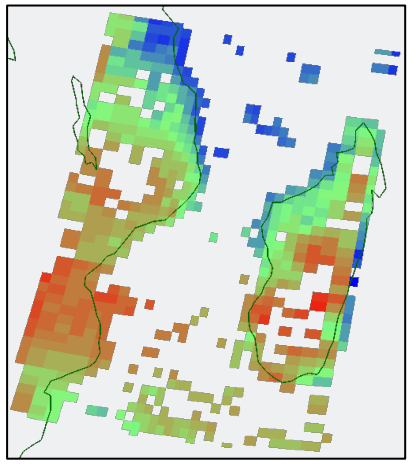


Mask based on user criteria (Quality level < 2)

Good quality data pixels retained



Rejected data points

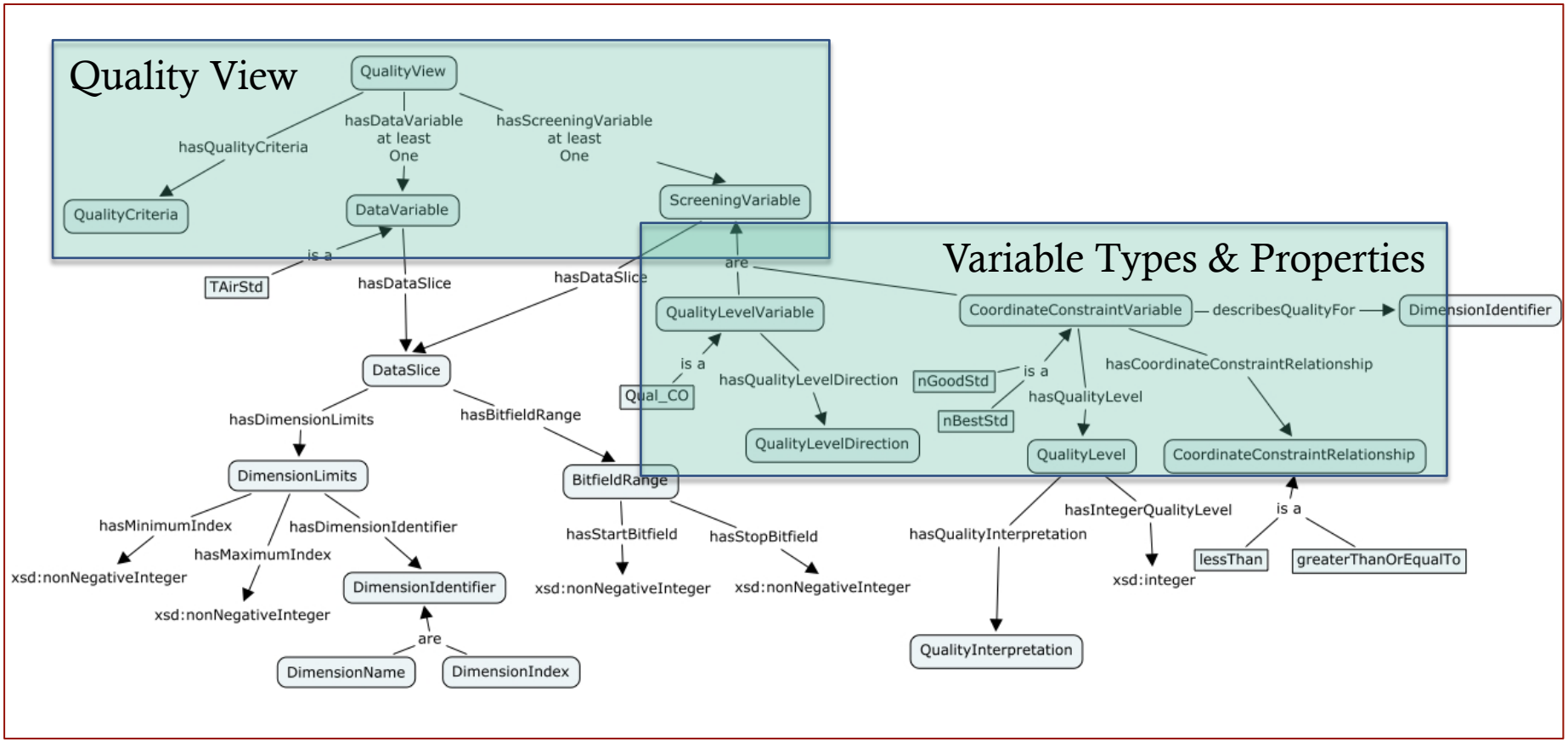


Output file has the same format and structure as the input file (except for the extra rejected-data fields)

DQSS encodes the science team recommendations on quality screening

- *Straightforward*: "Use Best-only for data assimilation uses; Use Best+Good for climatic studies"
- *More complicated*: "Use only VeryGood over land; Use Marginal+Good+VeryGood over ocean"

Ontology-driven design helps scale with diversity of quality schemes



Key Challenge: Modeling

- Science team recommendations for screening
- Uses of data
- Fitness for use
- Meaning/provenance of quality classifications