Semantic Web Data Discovery of Earth Science Data at NASA Goddard Earth Sciences Data and Information Services Center (GES DISC)
(http://mirador.gsfc.nasa.gov)

Mahabaleshwara Hegde[1,2], Richard F. Strub[1,2], Christopher S. Lynnes[1], Hongliang Fang[1,3], William Teng[1,3]
([1]NASA/GES DISC, [2]ADNET Systems, [3]Wyle Information Systems)

NASA Goddard Earth Sciences (GES)
Data and Information Services Center (DISC)
Christopher.Lynnes@nasa.gov
http://disc.sci.gsfc.nasa.gov/

Mirador is a web interface for searching Earth Science data archived at the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC). Mirador provides keyword-based search and guided navigation for providing efficient search and access to Earth Science data. Mirador employs the power of Google's universal search technology for fast metadata keyword searches, augmented by additional capabilities such as event searches (e.g., hurricanes), searches based on location gazetteer, and data services like format converters and data sub-setters. The objective of guided data navigation is to present users with multiple views (e.g., Projects, Earth Science Parameters, Applications) of all the available data in Mirador. Starting with any of the top level hierarchies, users can quickly navigate down to locate data of interest. The heart of the guided navigation in Mirador is an ontology based on the Global Change Master Directory (GCMD) Directory Interchange Format (DIF). Current implementation includes the project ontology covering various instruments and model data. Additional capabilities in the pipeline include Earth Science parameter and applications ontologies.

Data Discovery at NASA GES DISC

Data Holdings at NASA GES DISC

- Atmospheric Composition
  - Atmospheric Infrared Sounder (AIRS), Microwave Limb Sounder (MLS), High Resolution Dynamics Limb Sounder (HIRDLS), Ozone Monitoring Instrument (OMI), Total Ozone Mapping Spectrometer (TOMS), Solar Radiation and Climate Experiment (SORCE), Limb Infrared Monitor of the Stratosphere (LIMS) Upper Atmosphere Research Satellite (UARS), Microwave Sounding Unit (MSU)
- Hydrology
  - Global Land Data Assimilation System (GLDAS)
- A-Train
  - CloudSat-collocated MODIS/Aqua, OMI/Aura, POLDER data and MLS/Aura-collocated MODIS/Aqua data
- Precipitation
  - Tropical Rainfall Measuring Mission (TRMM)

Current Data Search Tool at NASA GES DISC

- Keyword search for data based on indexing of metadata using Google appliance
  - Useful in ranking matched data
- Spatial search based on representation of data's spatial foot print using rectangular mesh
  - Improved performance at the expense of accuracy
- Spatial search based on location gazetteer
  - Allows searching by geographical names
- Event search by storing spatial-temporal foot prints of events. Supported events include hurricanes, volcanic eruptions etc.
  - Allows searching for events by name (e.g., Katrina)

Keyword Search Interface	Matching Data Sets	Matching Data Files

Limits of Using Keyword Search in Data Discovery

- Some users are "navigation-dominant", not "search-dominant".
- Keyword search does not expose relationships among data.
- High precision in keyword search requires some prior knowledge of data.
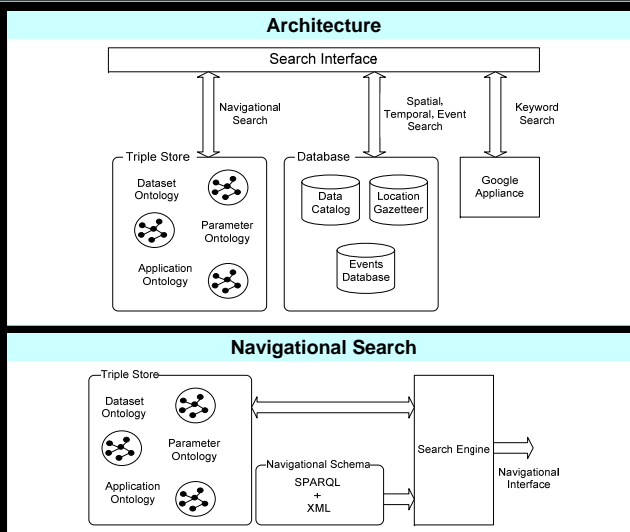
Application of Semantic Technology

Challenges in Searching Earth Science Data

- Data set may have their own set of unique metadata (e.g., processing level, data format, projection, etc)
- Projects and data producers wish to classify data in their own preferred way (e.g., grouping based on processing level, based on data format, etc)
- Some parameters may have different definitions depending on how they are measured and processed (e.g., total ozone)
- Parameters can have aliases (e.g., rainfall, total precipitation)

The Promise of Semantic Technology

- Resource Definition Framework (RDF) data store, the building block of Semantic technology, is flexible and extensible.
- Flexibility of RDF leads to easier solutions for assembling and managing large amounts of information and relating them to one another.
- Key technologies that make up Semantic technology, RDF, RDFS and OWL, provide myriad ways of representing relationships in data.
- The schema information for RDF, RDFS, is stored in RDF making it possible to query the schema using same tools used in querying the data store.

Architecture

Navigational Search

Data Navigation Using Semantic Technology

Sample Dataset Ontology
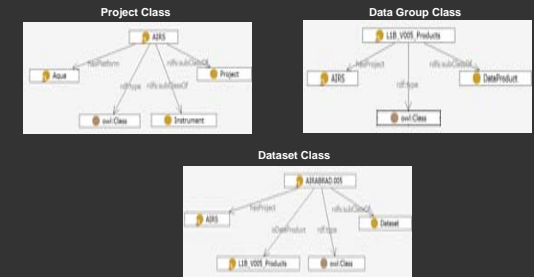
Navigation by Project	Navigation by Data Group
Navigation by Dataset	Matching Data Files

Work in Progress

- Navigation by parameters using an ontology associating parameters with data
- Navigation by applications using an ontology associating parameters with relevant applications