Integrating atmospheric and volcanic gas data in support of climate impact studies using semantic technologies

Peter Fox (HAO/ESSL/NCAR)
Krishna Sinha (VT), Rob Raskin (JPL), Deborah McGuinness (RPI)
Funded by NASA/ACCESS and NASA/ESTO/AIST
**SESDI Impact:** A Better Way to Access Data

**The Problem**

Scientists only use data from a single instrument because it is difficult to access, process, and understand data from multiple instruments. A typical data query might be:

- “Give me the temperature, pressure, and water vapor from the AIRS instrument from Jan 2005 to Jan 2008”
- “Search for MLS/Aura Level 2, SO2 Slant Column Density from 2/1/2007”

**A Solution**

Using a simple process, SESDI allows data from various sources to be registered in an ontology so that it can be easily accessed and understood. Scientists can use only the ontology components that relate to their data. An SESDI query might look like:

- “Show all areas in California where sulfur dioxide (SO2) levels were above normal between Jan 2000 and Jan 2007”

This query will pull data from all available sources registered in the ontology and allow seamless data fusion. Because the query is measurement related, scientists do not need to understand the details of the instruments and data types.
Volcano–Atmosphere considerations

- Focus on tropopause -> temperature gradients
- Stratospheric and tropospheric aerosols, the tropospheric reservoir
- Quantities/processes: Gas, particles, ejecta, scattering
- Records: Pulses, e.g. in SO2 events
- Related aspects: SO2, H2SO4, O3 chemistry
- Data from: in-situ and remotely sensed observation, proxy, simulation, pseudo-proxy
- Processes: solar, volcanic, GHG, ocean, land-use
- Priors to consider: statistics of variability and extremes
- Main task: detection and attribution
- Solar–Atmosphere considerations are very similar
Components to implement

- An analysis application
- Cross-domain terms, concepts and relations
- Connections to underlying data
Detection and attribution relations...
Modified GEON Solution Framework

Data Discovery

Level 1:
Data Registration at the Discovery Level, e.g. Volcano location and activity

Level 2:
Data Registration at the Inventory Level, e.g. list of datasets by types, times, products

Level 3:
Data Registration at the Item Detail Level, e.g. access to individual quantities

Earth Sciences Virtual Database
A Data Warehouse where Schema heterogeneity problem is Solved; schema based integration

Data Integration

Ontology based Data Integration
How to find the data?

• Think about it the way the data providers do.
SEDRE: Semantically Enabled Data Registration Engine

- SEDRE: a system that enables scientists to semantically register data sets for optimal querying and semantic integration
- SEDRE enables mapping of heterogeneous data to concepts in domain ontologies
Semantic Registration in SEDRE: An Overview

- SEDRE is a desktop application
- Users download and install SEDRE
- SEDRE accesses domain ontologies
- Users map data attributes (e.g., SO$_2$) to concepts in ontologies without ‘knowing it’
Example 1: Registration of Volcanic Data

Location Codes:
- **U** - Above the 180° turn at Holei Pali (upper Chain of Craters Road)
- **L** - Below Holei Pali (lower Chain of Craters Road)
- **UL** - Individual traverses were made both above and below the 180° turn at Holei Pali
- **H** - Highway 11

$\text{SO}_2$ Emission from Kilauea east rift zone - vehicle-based (Source: HVO)

Abreviations: t/d=metric tonne (1000 kg)/day, SD=standard deviation, WS=wind speed, WD=wind direction east of true north, N=number of traverses
Loading Volcanic Data into SEDRE
Registering Volcanic Data (1)
Registering Volcanic Data (2)

- No explicit lat/long data
- Volcano identified by name
- Volcano ontology framework will link name to location
Example 2: Registration of Atmospheric Data

Satellite data for $\text{SO}_2$ emissions

Abbreviation: SCD: Slant Column Density (in Dobson Unit (DU))
Loading Atmospheric Data into SEDRE
Registering Atmospheric Data (1)
Registering Atmospheric Data (2)

The image shows a user interface for registering atmospheric data. The interface includes sections for location (latitude, longitude, elevation), major elements (SO₂, HCl, H₂SO₄, H₂O, HNO₃, CO, N₂O, CION₂), and compounds (sulfides, sulfates, bromides, iodides, acetates, thiocyanates, nitrides, oxybromides, oxychlorides, oxyfluorides, sulfuric acid). The interface also includes a table for data attributes and acquisition details, such as latitude, longitude, and SO₂ concentration (SCD). The data is part of a NASA initiative for atmospheric data registration.
SEDRE+DIA: Overview

DIA: Web-based System for Data Discovery, Integration and Analysis

(Developed at Virginia Tech through NSF funding)
Summary

- Semantic data frameworks technologies are changing the landscape of providing data to scientists.
- Tools for data registration are soon to be available.
- Applications to perform data integration mediated by semantics are available.
- Initial results – applied to two volcanoes – led to correlation of SO2 concentration from volcano and in the atmosphere and relation to H2SO4.
Semantic Web Methodology and Technology Development Process

- Establish and improve a well-defined methodology **vision** for Semantic Technology based application development
- Leverage any existing vocabularies

**Use Case**
Small Team, mixed skills

**Analysis**
Develop model/ontology

**Open World:**
Evolve, Iterate, Redesign, Redeploy

**Rapid Prototype**
Leverage Technology Infrastructure

**Adopt Technology Approach**
Science/Expert Review & Iteration

**Use Tools**
Atmosphere Use Case

- Determine the statistical signatures of both volcanic and solar forcings on the height of the tropopause
Procedure

- Decompose use case vocabulary (nouns and verbs) for concepts and relations
- Put aside processes – treat them axiomatically
- Develop ontologies using concept mapping tool
- Leverage data infrastructure and registration procedures
- Exercise the use case
SWEET 2.0 Modular Design

- Supports easy extension by domain specialists

- Organized by subject (theoretical to applied)

- Reorganization of classes, but no significant changes to content

- Importation is unidirectional

Math, Time, Space
Basic Science
Geoscience
Processes
Geophysical Phenomena
Applications

importation
SWEET 2.0α Ontologies

Math
- Statistics
- Geometry
- Vector
- Units
- Space
- Time

Informatics
- Data
- System

Science
- Biology
  - Plant
  - Animal
- Ecology
- Physics
  - Mechanics
  - Thermo
  - ElecMag
- Nuclear
  - Waves
- Radiative Transfer
- Gephys
  - Fluid Dyn

Chemistry
- Matter
- Element
- Compound

SocialSci

PlanetaryScience
- SolidEarth
- Tectonics
- Volcano
- Seismo
- Bound Layer
- Cloud
- Front
- Precip
- Electric
- Atmosphere
- Dynamics
- Circulation
- Ocean
- Floor

Cryosphere
- Heliosphere
- Sun
- Surface
- Soil
- Landscape
- Coast

Applications
- Hazards
- Hydrology
- Agriculture
- Infrastructure
- Admin Borders
Registration of Data

• **Level 1.** Discovery of data resources (e.g., by discipline/sub-discipline – data type) requires registration through use of high level index terms.

• **Level 2.** Discovering Item level databases requires registration at **data level/type ontologies** (e.g. geochemistry, gravity atmospheric chemistry database).

• **Level 3.** Item detail level registration with independent variables, etc. (e.g., lat, lon, time, column in geochemical database that represents SO2 measurement). This level of registration is a requirement for semantic integration.