

# An Overview of VOs for Integrating Space and Solar Physics Data

M. Weiss, *JHU/APL*

D. A. Roberts, *NASA GSFC*

# Overview

- Heliophysics infrastructure and how VOs fit
- Current HP VOs
- New NASA HP VOs and related efforts
- SPASE Data Model as an integrator
- Other cross-VO issues
- Actual and potential services

## Definition of a “Virtual Observatory”

A Virtual Observatory (VO) is a **suite of software applications** on a set of computers that allows users to **uniformly find, access, and use resources** (data, software, document, and image products and services using these) from a collection of **distributed product repositories** and **service providers**.

A VO is a service that unites services and/or multiple repositories.

# The HP Data Environment

- Data from the [Heliophysics Great Observatory](#) reside in a distributed environment and are served from multiple sources.
- Multimission Data Centers
  - Solar Data Analysis Center
  - Space Physics Data Facility (CDAWeb, OMNIWeb, etc.)
  - National Space Science Data Center
  - NGDC/SPIDR
  - CDPP
  - DARTS ...
- Mission-level active archives: e.g. ACE, TIMED, TRACE, Cluster, Ulysses, etc.
- Much of our data are served from individual instrument sites.
- We are moving into a new data environment of:
  - Virtual Observatories for convenient search and access of the distributed data, and
  - Resident Archives to retain the distributed data sources even after mission termination (NASA).

# NASA Heliophysics Mission Data Lifecycle

Event	Launch	End of Prime	End of New Data	End of Mission	Final Archiving
Data Actions	Data system production	Reduction, PI serving, product improvement		Final data file prep., PI Serving	Serving; PI help Preserving Serving
Archive Status		Active Archive (mission-based; others possible in all phases)			Resident Archive Permanent Archive
Data Mgmt Document	PDMP Prep. and approval	PDMP use; MAP preparation and use in the Extended mission			RA plan Archive Plan (MOU)
General Servicing		Virtual Observatories and Services			
Management	Mission (Program/Project)		Mission (Prog./Proj.) (with Senior Review)		NSSDC

# Basic Philosophy

- Evolve the existing HP data environment:
  - take advantage of new computer and Internet technologies to
  - respond to our evolving mission set and community research needs (enable the HP Great Observatory)
- Blend ‘bottoms-up’, ‘market-driven’ implementation approaches with a ‘top-down’ vision for an integrated data environment.
- Assure that the HP science community participates in all levels of data management.

## (eGY) Principles Essential to VO Success

- All data produced by the HP missions will be open and made available as soon as is practical.
- Data will be independently scientifically usable. This requires:
  - adequate documentation including uniform (e.g., “SPASE”) descriptions
  - easy electronic access
  - sustainable and open data formats
  - provision of appropriate analysis tools.

## Architecture of the generic “Virtual HP Great Observatory”

- The environment will remain distributed
  - Many archives with different internal workings
- Data integration capabilities provided by discipline-based (“VxO”) and other more general virtual observatories
  - Linked by a central dictionary (“SPASE Data model” at least for NASA and some related VOs) as well as machine-to-machine interfaces.
  - Easily permits the inclusion of essential data sets from multiple sources across agencies and countries.
  - Provides a context for services and advanced analysis tools

# Other “Interoperability” Issues

- The following are desirable:
  - Registry services that allow posting and harvesting of resource descriptions
  - Tools for creating metadata to describe resources
    - requires uniform data model to work for all
  - A common API for accessing VOs and providers
    - e.g., a “SPASE query language”
- Some of these exist in other communities (e.g., NVO/IVOA), but HP has a different and in some ways more complex data model.
- There is an increasing need for interagency and international cooperation (again, as in other communities)

# Community Approach

- Use peer-review processes to assist in managing the elements of the environment.
  - Proposals for, e.g.:
    - VOs
    - Data restoration and access improvement
    - Resident Archives
    - Value-added services
  - Periodic reviews of VOs, Data Centers, etc.
- Success will be determined by community use and feedback. The process is “market-driven.”

# Early Heliophysics VOs (~2001)

- VSO (Virtual Solar Observatory)
  - A “small box” that delivers solar physics images, spectra, etc. as FITS files or thumbnails; covers most solar physics products
  - Recently added browse movies, searching by event lists, a shopping cart, and other features
  - Path-finder for “VxOs”
- VSPO (Virtual Space Physics Observatory)
  - A “large box” that provides efficient access to a wide variety of space and solar physics products, including data from over 100 observatories/spacecraft
  - Links to other VOs, etc. (VSO, VHO, CDAWeb, SSCWeb, ftpbrowser) for delivery of data and services
  - Uses and delivers SPASE metadata for resources

# Early Heliophysics VOs

- EGSO (European Grid of Solar Observations; to become “HELIO”; EU/ESA)
  - Emphasizes using catalogued events and features to constrain searches
  - Primarily solar initially, but expanding to all of HP
- (VO related early work) CDAWeb and related services available through an API that allows machine access to data.
- (Other fields) NVO (astronomy), OPeNDAP (Earth science)
- Many basic ideas go back to a “Space Science Data System” envisioned over 20 years ago.

# Newer NASA HP VxOs

- VHO (Heliosphere; GSFC)
  - Has a prototype that delivers some data and allows multiple query types; working on data descriptions, and collaborating with VMO-G to enhance middleware.
- VITMO (Ionosphere, Thermosphere, Mesosphere; APL)
  - Existing prototype delivers data based on parameter- and event-based searches; will incorporate a coincidence tool to aid searches.
- VMO (Magnetosphere; GSFC and UCLA)
  - Developing many SPASE descriptions as well as useful registry tools and middleware.
- ViRBO (Radiation Belts; GMU)
  - Currently ingesting datasets and developing data descriptions and a web portal; planning a strong connection to models

# Related Heliophysics VOs

- GAIA (Global Auroral Imaging and Access; Canada)
  - Integrates all-sky camera data from around the world; serves increasing amounts and variety of ground-based HP data
- VSTO (Virtual Solar-Terrestrial Observatory; NSF/NCAR)
  - Portal for CEDAR (ground-based) data; also MLSO (solar); broad charter.
  - Based on ontologies and related tools.
- STARS (Solar-Terrestrial data Analysis and Reference System; Japan)
  - Builds on the DARTS system; currently PC (Windows) based.

# VO “Glue”: Metadata and Data Models

- Uniform access requires uniform metadata description: i.e. a “data model” or “ontology”
- Metadata about the data, written in the uniform language (e.g., “SPASE”) is essential (although tedious to produce)
- A VO is only as good as its metadata
- SPASE is providing the data model for NASA HP VOs; we may end up needing translators.
- Individual VOs and provider internals may have different metadata, but should translate to a common language for interoperability

# Services

- The most essential services are **searching** for relevant data using data model terms, and **retrieval** of data files.
- The next most common services are **subsetting**; providing **translated** (e.g. ASCII) files; and **plotting/imaging** data in a “browse” mode through on-the-fly plots or stored GIFs/JPEGs.
- General **format translation** tools exist, although not yet integrated into HP VOs; general **coordinate system translation** is possible, but also not yet implemented.
- Other useful services include **visual data selection** tools, and automated data **interpolation** and **rebinning**.
- “**Higher order queries**,” ultimately involving natural language searches and associated event lists, are beginning to be implemented.
- Stronger **links between data and models** are essential long-term, but only beginning to be explored.

# Summary and Directions

- Heliophysics has made a good start in developing VOs for uniform data access and attendant services.
- Existing VOs in HP and other communities show that the basic ideas of distributed systems and “web service” tools will work in this field.
- Much work needs to be done to develop the basic metadata for the many existing resources.
- There is an increasingly urgent need for cooperation and collaboration between diverse efforts.
- Long-term challenges include developing more powerful query tools and integrating data and model services.

See: <http://hpde.gsfc.nasa.gov> for more documents, links, and details.