Virtual Observatory: A Tool to Support Data Access & Data-Model Comparison

Shing F. Fung
Geospace Physics Laboratory
NASA Goddard Space Flight Center

CCMC Workshop, Annapolis, Maryland, April 11-15, 2016
Accessing Data from NASA Heliophysics Archives

Distributed data sources

Different solar mission data
Heliospheric mission data
Magnetospheric mission data
ITM mission data

Data archives

SDAC

SPDF

Solar data users
Non-solar data users
Heliophysics VxOs: A New Paradigm

VxOs

- VSO
- VHO
- VMO
- VITMO

Distributed data sources

- Different solar mission data
- Heliospheric mission data
- Magnetospheric mission data
- ITM mission data

Data archives

- SDAC
- SPDF

Cross-disciplinary VxOs:

- VWO
- VEPO
- VIRBO
- VMR

Discipline-oriented users

- Solar data users
- Non-solar data users
ISWI current projects are 17 (January, 2015)

Numerous ground-based instruments (e.g., ISWI network) provide data that complement space-based measurements.
“Big Data” Is Coming!

• Instruments are becoming data-intensive, e.g.,
  – Solar Dynamics Observatory ~1.4TB/day (science telemetry)
  – MMS ~137GB/day (higher-level science)
  – 100’s-1000’s ground stations, generating ~GB/day/station

• Diverse [space, ground, model (e.g. CCMC)] & distributed (national & international) data sources

• Effective data services need to locate/search/access/retrieve/deliver data efficiently
  – Large data volume in distributed data sources
  – Different data types and products
  – Diverse user needs
General Heliophysics Information Flow

- Data archives/VxOs
- Space- & ground-based heliophysics data sources
- Retrospective data
- Science analysis
- Theory / Model development
- Validation / transition to operations
- Model data
- Test data
- Research

Flow Diagram:
- Data archives/VxOs to Science analysis
- Science analysis to Theory / Model development
- Theory / Model development to Validation / transition to operations
- Validation / transition to operations back to Data archives/VxOs
General Heliophysics Information Flow

Data archives/VxOs

Science analysis

Theory / Model development

Validation / transition to operations

“O2R”

“R2O”

Users

Space weather nowcast / forecast

Space & ground-based heliophysics data sources

Real-time data

Research

Operations

Retrospective data

Model data

Test data

Feedback to models
General Heliophysics Information Flow

- Science analysis
- Theory / Model development
- Validation / transition to operations
- Users / Data source
- Space weather nowcast / forecast
- Data services interface
- Space- & ground-based heliophysics data sources
- Retrospective data
- Test data
- Model data
- Real-time data
- Feedback to models

Data-model comparison interface

“O2R”

“R2O”
Virtual Observatory Middleware View

Users (individuals/data-use services)

VO Query Builder interface

Search engine applies input in search algorithms to locate data

VO API to access/retrieve data

Input: Metadata as query conditions (platforms, measurement type, time ranges, etc.)

Metadata & data delivery to users

VO API

Local data archive

Remote data centers

Distributed data sources

Modeling centers (e.g., CCMC)
Virtual Observatory Middleware View

Users
(individuals/data-use services)

Input: Metadata as query conditions (platforms, measurement type, time ranges, etc.)

VO Query Builder interface

Search engine applies input in search algorithms to locate data

VO API to access/retrieve data

Metadata & data delivery to users

Local data archive

Remote data centers

Distributed data sources

Modeling centers (e.g., CCMC)

Use of metadata Standards (e.g., SPASE) can effectively remove distinction between data sources
Summary

• A virtual observatory can provide enhanced data services to
  – Augment existing data archive services
  – Support Diverse & distributed sources (space, ground, model)
  – Support diverse data users (research & operations)
  – Support effective data search mechanisms
    • Discipline orientation
    • Targeted data volume

• Enabling technologies
  – Standard metadata model to reduce s/w complexity
    • Same tools to search & access multi-disciplinary data
  – Standard access protocols (APIs)
    • Ensure interoperability
    • Enable system expandability by interfacing with new services