SPASE 2.0: Standardization of Space Physics Data Access and Retrieval

Jim Thieman[1], Todd King[2], Aaron Roberts[3], et al.

[1] NASA Goddard Space Flight Center, Code 690.1, Greenbelt, MD, United States
[2] Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA, United States
[3] NASA Goddard Space Flight Center, Code 672, Greenbelt, MD, United States

OUTLINE

- The Nature of SPASE
- Version 2.0 of the SPASE Data Model
- Virtual Observatories and the Heliophysics Data Environment
- Facilitation of Data Model Applications
- The Future
- Conclusions

What is SPASE?

- Acronym for:
  Spase Physics Archive Search and Extract
- An organization to set community-based standards with the goals of:
  - Defining a data model for Space Physics
  - Demonstrating its viability
  - Enabling interoperability in a federated environment

So that...

- Resources can be easily registered, found, accessed, and used

A Brief History

Planning

- **1998 - ISTP**
  - The SPASE effort has its root in the data handling session of the ISTP workshop held at RAL in 1998, when on Sept 26 a resolution was passed calling on the "larger data centers" to "do something" to make data more accessible.
- **2001 - AISRP**
  - Early in 2001 a breadboard interoperability test bed was implemented between NSSDC and CDPP/CNES, and later that year, in response to an AO from NASA AISRP ROSS (Applied Information Systems Research Program, Research Opportunities in Space Science), a proposal entitled "A Space Physics Archive Search Engine (SPASE)" was submitted jointly by NSSDC, SwRI, RAL and CDPP.
- **2002 - Grassroots**
  - While this proposal was not funded a volunteer effort continued and attracted broader participation. It was recognized that a data model was needed to establish an "interlingua" to share resources across the entire space physics domain. The goals of this effort were defined in late 2002 and the new moniker of Space Physics Archive Search and Extract (SPASE) was adopted.

Design & Implement

- **2003 - Open Community – NASA LWS**
  - In 2003 the effort was organized as an international consortium with an open invitation for anyone in the community to participate. U.S. participants in SPASE were funded by NASA in July 2005 which helped accelerate the effort.
- **2005 - Release 1.0**
  - In November 2005 SPASE released version 1.0 of its ontology (data model).
- **2006 - Release 1.1.0**
  - In response to community feedback, the data model was improved and in August 2006 version 1.1.0 was released. In that same year NASA solicited proposals to establish thematic virtual observatories for the heliophysics community and SPASE was adopted as the metadata standard to enable interoperability.
- **2007 – Release 1.2.0**
  - Based on feedback from the community and from the selected virtual observatories the data model was further refined and version 1.2.0 was released in May 2007.
- **2009 – Release 2.0.0**
  - After a period of use in NASA's VxOs the model was streamlined and enhanced to support a wider range of resources. Released April 2009.
- **2009 – Release 2.0.1**
  - Additions to support forward migration of existing resource descriptions. Released July2009.

NASA Establishes VO's

International Cast of Characters

Augsburg College
Mark Engebretson, engebret@augsburg.edu
Noel Petit, petit@augsburg.edu

California Institute of Technology (CalTech)
Andrew Davis, ad@srl.caltech.edu

Centre de Données de la Physique des Plasmas (CDPP)
Michel Gangloff, gangloff@cesr.fr
Christopher Harvey, christopher.harvey@cesr.fr
Claude Huc, claude.huc@cnes.fr

Istituto Nazionale di Astrofisica (INAF)
Kevin Reardon, kreardon@arcetri.astro.it

Japan Aerospace eXploration Agency (JAXA) - STP/Ehime
Yasumasa Kasaba, kasaba@stp.isas.jaxa.jp
Ken T. Murata, STP/Ehime, murata@cite.ehime-u.ac.jp

Jet Propulsion Laboratory (JPL)
Dan Crichton, dan.crichton@jpl.nasa.gov
Steven Hughes, j.steven.hughes@jpl.nasa.gov

John Hopkins University/Applied Physics Laboratory (JHU/APL)
Rose Daley, rose.daley@jhuapl.edu
Brand Fortner, brand.fortner@jhuapl.edu
Daniel Morrison, daniel.morrison@jhuapl.edu
Stu Nylund, stu.nylund@jhuapl.edu
Jon Vandergriff, jon.vandergriff@jhuapl.edu
Michele Weiss, michele.weiss@jhuapl.edu

George Mason University
Robert Weigel, rweigel@gmu.edu

Goddard Space Flight Center (GSFC)
Ed Bell (PSGS), ed.bell@gsfc.nasa.gov
Dieter Bilitza (RITSS), bilitza@mail630.gsfc.nasa.gov
Bobby Candey, candey@mail630.gsfc.nasa.gov
Carl Cornwell (Aquilent), carl.cornwell@aquilent.com
Joe Gurman, gurman@grace.nascom.nasa.gov
Joe Hourcle (EITI), oneiros@grace.nascom.nasa.gov
Mona Kessel, kessel@ndadsb-f.gsfc.nasa.gov
Joe King (PSGS), jking@mail630.gsfc.nasa.gov
Terry Kucera, kucera@stars.gsfc.nasa.gov
Bob McGuire, rmcguire@pop600.gsfc.nasa.gov
Jan Merka, jan.merka@gsfc.nasa.gov
Lou Reich (CSC), lreich@pop500.gsfc.nasa.gov
Aaron Roberts, roberts@vayu.gsfc.nasa.gov
Don Sawyer, donald.sawyer@gsfc.nasa.gov
Dave Sibeck dsibeck@pop600.gsfc.nasa.gov
Adam Szabo, aszabo@pop600.gsfc.nasa.gov
Jim Thieman, james.r.thieman@nasa.gov
Karen North, Karen.C.North@nasa.gov
Aaron Smith (Aquilent), aaron.smith@aquilent.com
Isaac Verghese (Aquilent), Isaac.Verghese@aquilent.com
Vasili Rezapkin (Aquilent), vasili.rezapkin@aquilent.com

National Aeronautics and Space Administration (NASA) HQ
Joe Bredekamp, jbredeka@mail.hq.nasa.gov
Chuck Holmes, cholmes@mail.hq.nasa.gov

National Oceanic and Atmospheric Administration (NOAA)
Eric Kihn, eric.a.kihn@noaa.gov

Rutherford Appleton Laboratory (RAL)
Chris Perry, c.h.perry@rl.ac.uk
Phil Richards, P.J.Richards@rl.ac.uk

Stanford University
Rick Bogart, rbogart@stanford.edu

Southwest Research Institute (SwRI)
Joey Mukherjee, jmukherjee@swri.org
Dave Winningham, david@cluster.space.swri.edu

University of California, Los Angeles (UCLA)
Steven Joy, sjoy@igpp.ucla.edu
Todd King, tking@igpp.ucla.edu
Ray Walker, rwalker@igpp.ucla.edu
Lee Bargatze, lfb@igpp.ucla.edu

which took…

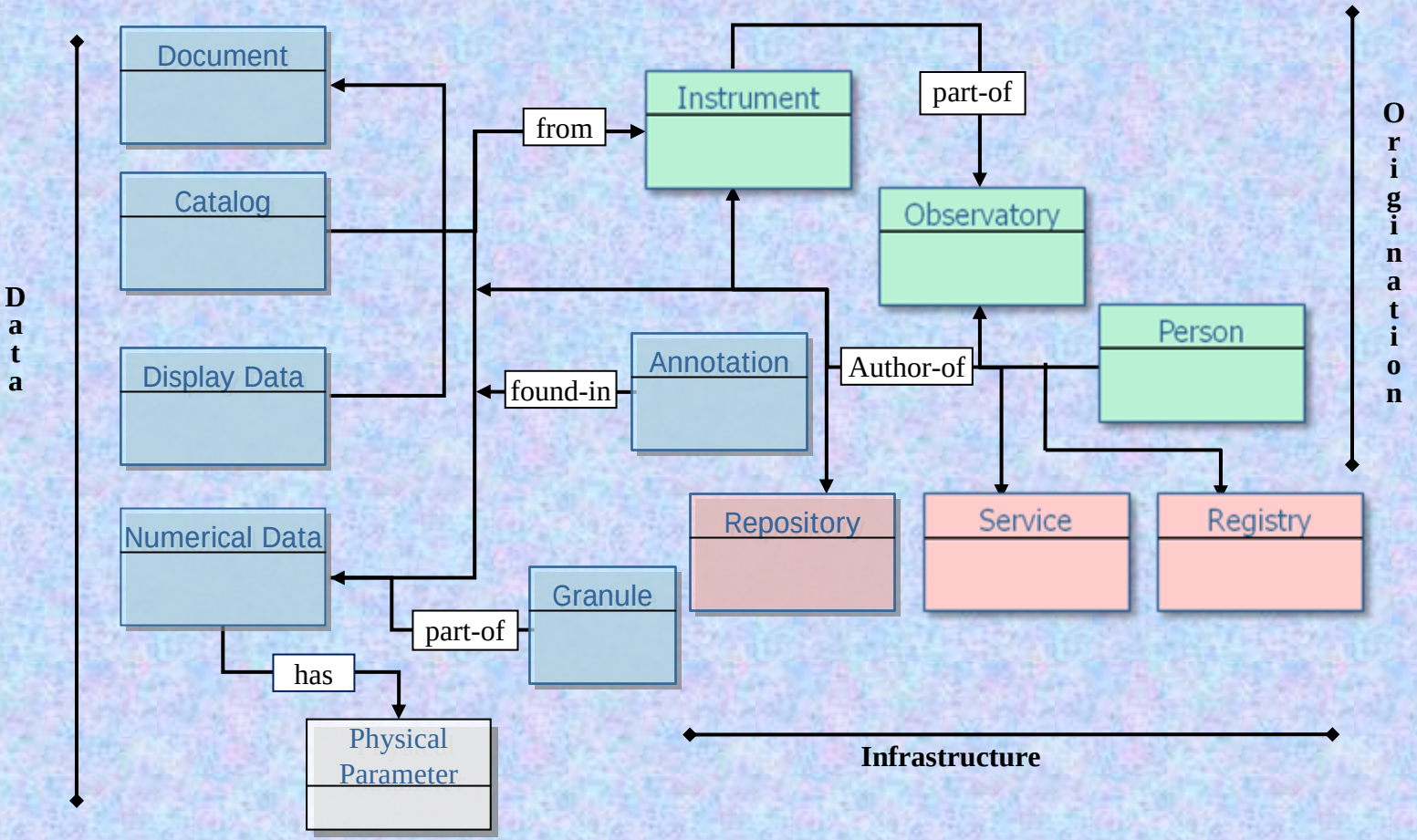
- Thousands of e-mails (4000+ since 2002)
- More than a hundred telecons (bi-weekly)
- A half-dozen face-to-face meetings

to achieve…

SPASE Today

- Data Model
  - Defined a standard data model for Heliophysics
    Current release version 2.0.2 (September 2009)
  - Vetted by research communities in the domains:
    - Magnetospheres
    - Waves
    - Ionosphere-Thermosphere-Mesosphere
    - Radiation Belts
    - Energetic Particles
    - Solar Physics
    - Models and Simulations
- Services
  - Initial work on metadata sharing (registries).
  - Distributed queries (SPASE-QL)

SPASE Conceptual Model



A simplified conceptual model of resources (classes) in the SPASE data model. Arrows point in the direction of association. Cardinality is not shown.

SPASE Speak

- The SPASE Data Model was developed by a team of scientists, IT specialists, data engineers and developers.
- Most documentation is oriented towards the user, so our nomenclature is less formal than the Unified Modeling Language (UML).

What we Say	In UML
Resource	Class
Element	Attribute
Container	Component Class
Association	Association

SPASE and ISO-11179

SPASE adheres to the principals of ISO-11179 (Metadata Registry) standard

- without formally adopting all aspects.

SPASE has a

- Data Element Concept (DEC) (part 1)
- Representation (part 1)
  - Value Domain
    - Enumerated
    - Non-enumerated
    - Value meaning
- Data Element Relationships (part 1)
- Data Element Formulation Rules (part 4)
- Naming and identification principles (part 5)
- Maintains a simple metadata registry (part3)

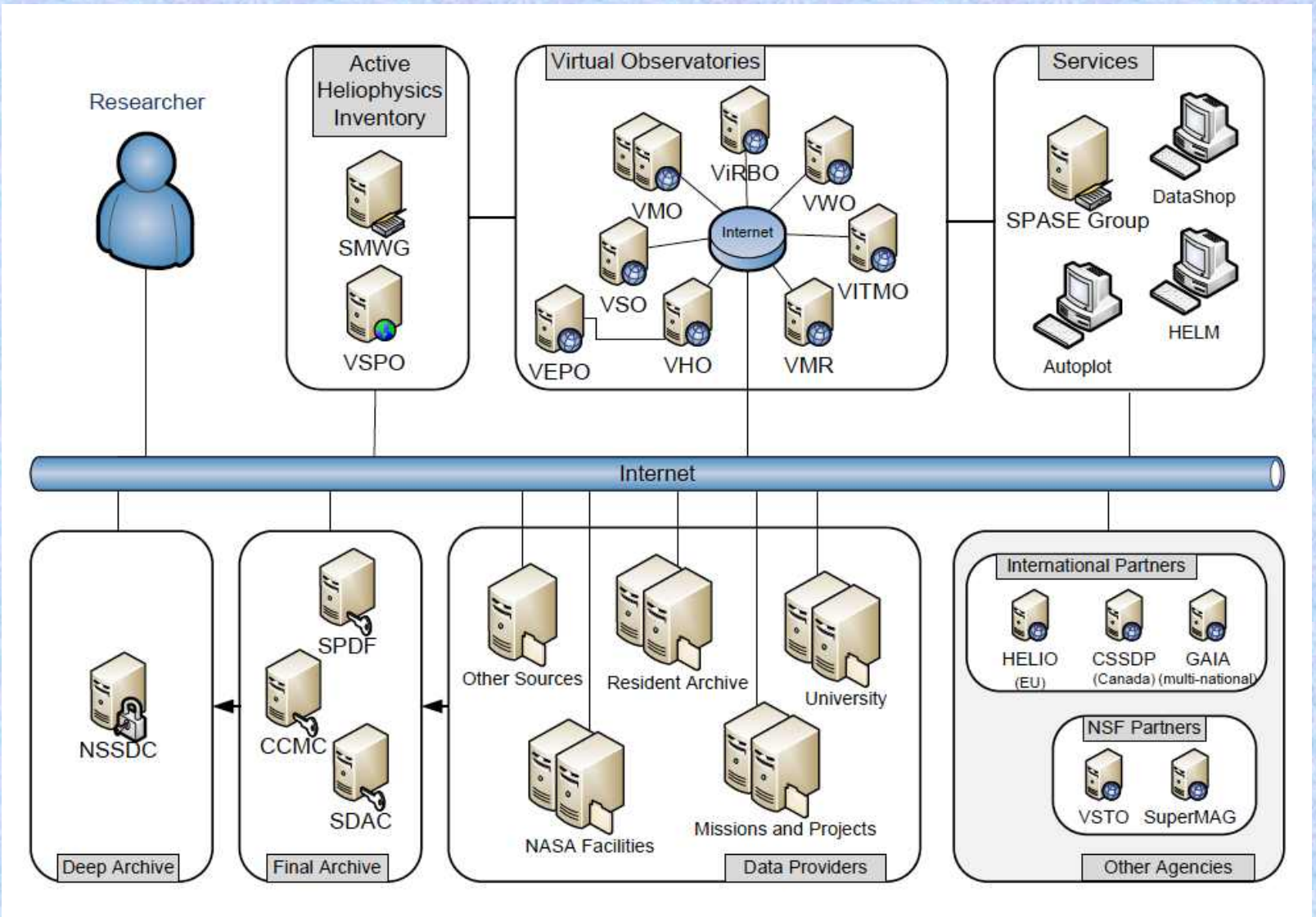
Data Model Design Principals

- **Data are self-documented.**
  Data resources have internal schema or structures for storing values.
- **Resources are distributed.**
  There are many providers of resources and these providers can be located anywhere in the world.
- **Online Resources have Universal Resource Locators (URL)**
  If a resource is on-line it can be accessed and retrieved using Universal Resource Locators (URL).
- **The data environment is continually evolving.**
  New resources are actively generated either as part of an on-going experiment or as a result of analysis and assessment.

Additional Details

- The SPASE Data Model is a Semantic Data Model.
  - Defines the meaning of data within the context of its interrelationships with other data.
  - Defines the scientific context of data.
- ...with Ontology features
  - Typed associations between resources.
- The SPASE Data Model is implementation neutral.
- Chosen reference implementation is XML.
  - XML Schema
  - Numerous XML style sheets for converting metadata.
    - To HTML
    - To OAI
  - And lots of tools.

Heliophysics Data Environment



Heliophysics Virtual Observatories (VOs)

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

NASA-Funded

- VSO - Virtual Solar Observatory
- VSPO - Virtual Space Physics Observatory
- VMO - Virtual Magnetospheric Observatory
- VITMO - Virtual Ionosphere, Thermosphere, Mesosphere Observatory
- VHO - Virtual Heliospheric Observatory
- ViRBO - Virtual Radiation Belt Observatory
- VEPO - Virtual Energetic Particle Observatory
- VWO - Virtual Wave Observatory
- VMR - Virtual Model Repository

Non-NASA-Funded

- CAA - Cluster Active Archive
- CDPP - Centre de Données de la Physique des Plasmas
- CSSDP - Canadian Space Science Data Portal
- EGSO - European Grid of Solar Observations
- GAIA - Global Auroral Imaging Access
- VSTO - Virtual Solar Terrestrial Observatory
- ??
- ??

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

SPASE in the Wild

Clarity comes from usage

Current Users (to different degrees):

- United States
  - NASA's Heliophysics Data Environment (HPDE)
    - 9 Virtual Observatories
    - Service providers (Autoplot, HELM, etc.)
  - National Science Foundation (NSF)
    - SuperMAG project
- Canada
  - Canadian Space Science Data Portal (CSSDP)
- European Union
  - Cluster Active Archive (CAA)
  - HELIO

Application Tools

Tools for working with SPASE metadata and the SPASE framework.

Validator

Determines compliance with a version of the SPASE data model.

XML Validate

Parser

Convert SPASE XML to internal structures

Parser

Editor

Web-based Editors

Web Editor

Standalone Editors

SPASE Assistant

Editors with Database Storage

Web+DB Editor

Generator

Create SPASE descriptions using external sources of information

Ruleset Description Generator

Harvester

Extracts information from SPASE resource descriptions (or registries)

SPASE Registry Server

SPASE Database Query

Wrapper

Converts or embeds SPASE metadata in other descriptions or forms (i.e., OAI)

Data Dictionary Lookup

SPASE-to-OAI mapping

Correlator

Divide an XML document into individual resource descriptions into a well organized file system

Correlator

Refcheck

Determine the validity of all references in a resource descriptions. Checks Resource IDs and URL

Refcheck

There are additional tools in development:

SPASE Query Language

Java-to-XML Binding Mechanism (JAXB)

SPASE Guidelines Document

Example Person Resource (XML)

```
<?xml version="1.0" ?>
  <Spase xmlns="http://www.spase-group.org/data/schema">
  <Version>2.0.0</Version>
  <Person>
    <ResourceID>spase://SMWG/Person/Todd.King</ResourceID>
    <ReleaseDate>2007-06-07</ReleaseDate>
    <PersonName>Todd King</PersonName>
    <OrganizationName>UCLA/IGPP</OrganizationName>
    <Address>3846 Slichter Hall
             Los Angeles, CA
             90095-1567
    </Address>
  </Person>
  </Spase>
```

The Future – Data Model

- Provide improved support and services:
  - Data Model version migration
  - Improved editors
  - Improved style sheets
  - Registries
  - ...and more
- Documentation:
  - Tutorials
  - Guides
  - References

The Future – Interfaces

- Services
  - Resource Registry
  - Reporting
  - Visualization (Autoplot)
  - Domain search engine
- Query API:
  - SPASE-QL
  - REST

The Future – and You

- **Join the fun.**
- If you have resources ...

  data, event lists, anything

  ... you want to share contact one of the current systems to get started.

- If you have a lot of resources to share consider establishing a "personal" virtual observatory with a SPASE registry.

Conclusion

Domain specific data models and ontologies are making it possible for the seamless exchange of data across groups, agencies and international boundaries. With sufficient support (parsers, services, etc) adoption can be easy. With the right documentation mastering SPASE is possible to all. SPASE is working towards this goal.

For more details:

www.spase-group.org

ABSTRACT

SPASE stands for Space Physics Archive Search and Extract. Over a number of years the international SPASE Working Group (see http://spase-group.org) has developed the SPASE Data Model which is a standard for describing data sets in the Space and Solar Physics (now often called Heliophysics) domain. The Data Model is a set of terms and values along with the relationships between them that allow describing all the resources in a heliophysics data environment. It is the result of an effort to unify and improve on existing Space and Solar Physics data models. The intent of this Data Model is to provide the means to describe resources, most importantly scientifically useful data products, in a uniform way so they may be easily registered, found, accessed, and used.

SPASE released version 2.0 of its data model in April 2009. This version represents a significant change from the previous release. It includes the capability to describe a wider range of data products and to include expert annotations which can be associated with a resource. Additional improvements include an enhanced capability to describe resource associations and a more unified approach to describe data product content. Details of these and other improvements will be discussed.

The SPASE Data Model has been used to document a large range of data sets of interest to the space physics community. Several approaches may be used to access data through these descriptions. The main method of data discovery is through Virtual Observatories (VO's) that have been set up in the U.S. and internationally to provide access to heliophysics data of interest. These VO's were established in particular heliophysics subdisciplines of interest and cater to users doing research within that subdicipline. The VO's are tuned to assist the type of research specific to the given subdiscipline so they sometimes differ from each other in the way they describe and handle data internally. SPASE is needed as a tool to enable standard terminology and access methods to be used to find and retrieve data from throughout the heliophysics data environment. The application of this approach will be demonstrated if internet access is available for the presentation or indicated by illustrations if it is not. Examples descriptions of data resources and the formation of resource collections will also be presented.

Version 2.0 of the SPASE Data Model provides a solid foundation for future development. The Data Model contains many inherent capabilities which are yet to be realized. Potential applications are discussed to illustrate the range of capabilities. This includes how information contained in other data models can be mapped to SPASE concepts and migrated between paradigms. This is essential in the international heliophysics data environment as more countries become involved and its heterogeneity broadens. Nonetheless, we seek to demonstrate the applicability of the SPASE Data Model by working with other countries having heliophysics

International Effort

- CNES/CNRS Plasma Physics (CDPP) Data Archive
- NASA/Goddard Space Flight Center
- NOAA/National Geophysical Data Center
- Planetary Data System- UCLA Plasma Physics Interactions Node
- Rutherford Appleton Laboratory
- Southwest Research Institute
- Applied Physics Laboratory
- Jet Propulsion Laboratory
- Institute of Space and Astronautical Science (ISAS/JAXA)
- Augsburg College
- European Grid of Solar Observations (EGSO)

SPASE Tools

- To demonstrate the viability of the model and provide basic support for its adoption a set of tools have been developed to support the reference implementation (XML) :
  - **Validator** - Determines compliance with a version of the SPASE data model.
  - **Parser** – Convert SPASE XML to internal structures.
  - **Editor** – Create SPASE descriptions by hand.
  - **Generator** – Creates SPASE descriptions using external sources of information.
  - **Harvester** – Extracts information from SPASE resource descriptions (or registries).
  - **Wrapper** – Converts or embeds SPASE metadata in other descriptions or forms (i.e., OAI).
  - **Correlator** – Divide an XML document into individual resource descriptions into a well organized file system.
  - **Refcheck** – Determine the validity of all references in a resource descriptions. Checks Resource IDs and URL.