

The SPASE Data Model: Standard Metadata for Space Science Data Description

J. Thieman, T. King, A. Roberts,
J. King, C. Harvey, and P. Richards

Oct. 24, 2006

OUTLINE

- The Space Physics Data Environment
- What is SPASE?
- The SPASE Data Model
- Scenarios for Data Access
- Application Tools
- Future Plans

Space and Solar Physics Data



QuickTime™ and a
YUV420 codec decompressor
are needed to see this picture.

Some Recent Space and Solar Physics Missions



Space Physics Data Environment



- Thousands of relatively small datasets as well as large datasets
- Tens to hundreds of data centers or data providers (repositories)
- Very loose (if any) coupling of data centers
- Very diverse metadata terminology and data formats
- Preservation of the data often becomes the task of NSSDC (a federal obligation in U.S.)



What is SPASE?

Space Physics Archive Search and Extract (SPASE)

- An International, community-based organization with the goals of:
 - Facilitating data search and retrieval across the Space and Solar Physics data environment
 - Defining and maintaining a standard Data Model for Space and Solar Physics interoperability
 - Demonstrating the Model's viability
- Initiated in late 2002 as a discussion among space physics data archive representatives.
- Organized in 2003 as an international consortium with an open invitation for anyone in the community to participate
- U.S. participants funded by NASA in July 2005. Will receive "permanent support" from NSSDC after 2007.



SPASE Goals

- To conceptualize the domain of space physics data and resources.
- Provide a formal dictionary (set of representational terms) to describe space physics resources.
- Provide a standard method of describing resources.

So that we can...

- Enable searching for data among the diverse space physics data repositories
- Facilitate intercomparison of similar quantities from different data sets in different data centers through common terminology mapping (e.g. visible radiation vs. optical radiation)

SPASE Activities



- Assemble a team of domain experts (scientists), information specialists and technologists to advance the goal of establishing standards for sharing space physics resources.
- Establish a forum for the discussion of a **Space Physics Data Model and Data Dictionary**.
- Fully discuss (e-mail + bi-monthly telecon) each addition to the model.
- Meet regularly (twice a year) to reach a formal consensus and release stable versions of the data model.
- Define and support useful representational forms of the data model (XML) with emphasis on numeric data descriptions first
- Test data model and its representation (XML) with user scenarios and real world resources.
- Refine data model in response to community needs.
- Encourage and support adoption of data model by providing tools and reference implementation.

The SPASE Data Model



- Version 1.0.0 released in November 2005
- Version 1.1.0 released in August 2006
Resource Types: Catalog
Display Data
Numerical Data
Granule
Instrument
Observatory
Person
Registry
Repository
Service
- Work is underway for version 1.2.0 which will contain extensions based on community needs (probable release - mid 2007).

Data Model Access



SPASE Space Physics Archive Search and Extract
Space Physics Archive Search and Extract (SPASE) Consortium

Home
Steering Committee
Data Model Working Group
Technical Working Group
Tools and Services
Consortium Members

Announcements:
SPASE face-to-face meeting (November 8-10, 2006) [more...](#)

The SPASE data system is a model for scientific data systems. It is based on the latest web-based technologies and is designed to be a distributed data systems with a heterogenous mix of platforms and systems.

These pages focus on the data model for the SPASE data system. The data model includes the structure of messages passed between systems; how to enrich data for interchange and archiving; and a data dictionary defining all terms and keywords used in the system. A full description of the data model is included under [Documents](#).

Also included are [examples](#) that implement the data model.

[Tools](#) to demonstrate the utility and capability of the SPASE metadata and framework

If you should have any questions or comments please [contact](#) us.

The [members](#) of SPASE include representatives from the international community.

Data Model Document
[Current Version \(1.1.0\)](#)
[Current Draft \(generated\)](#)
[All documents](#)

Data Dictionary
[Search](#)
[Tree](#)
[XML Schema](#)
[XML Stylesheet](#)

News
[Briefs](#)
[RSS](#) [XML](#)

Documents
[Charters](#)
[Meetings](#)
[Presentations](#)
[Standards](#)

← Version 1.1.0 is now available

Available at:
<http://www.spase-group.org>

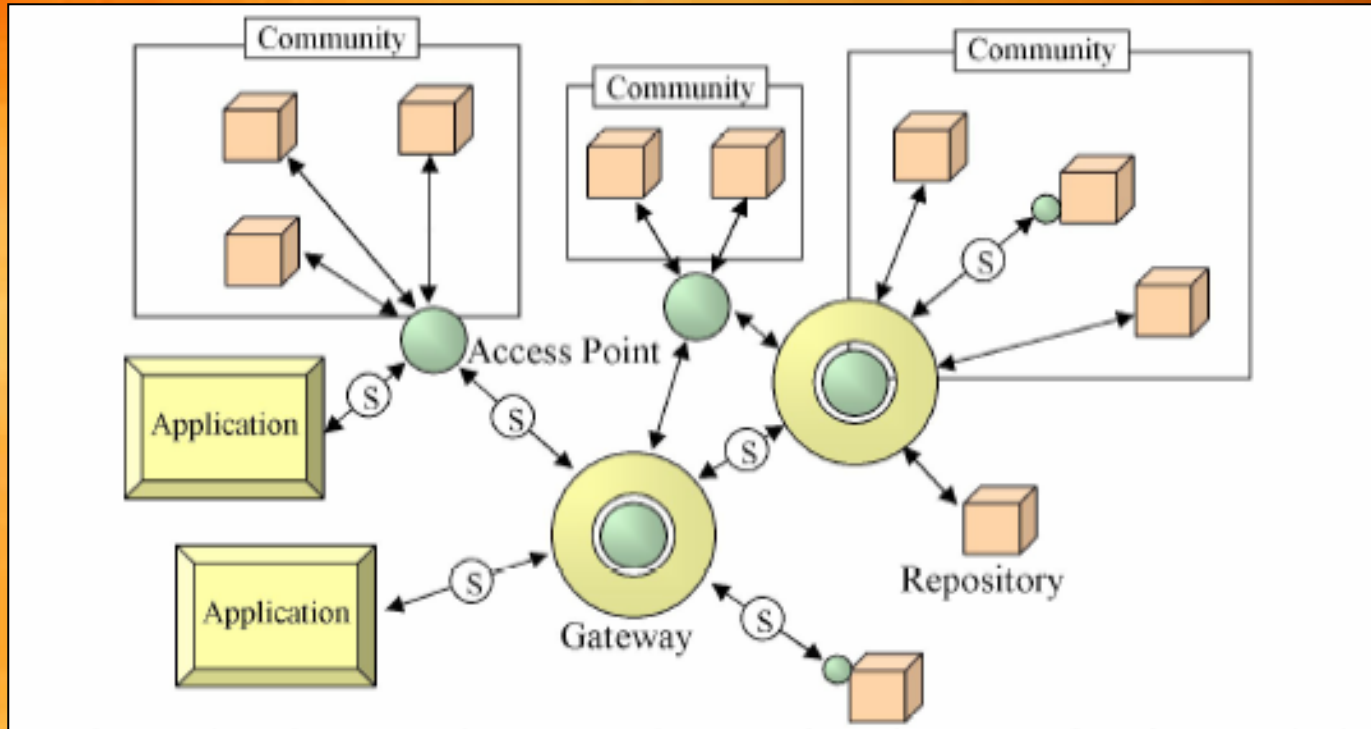


Current Adopters (or nearly so)

- NASA's Virtual Observatory effort
 - VSO (Solar)
 - VHO (Heliospheric)
 - ViRBO (Radiation Belt)
 - VITMO (ITM)
 - VMO/G (Magnetospheres)
 - VMO/U (Magnetospheres)
 - VSPO (Space Physics)
- ESA Cluster Data Archive
- Centre de Données de la Physique des Plasmas (CDPP)

All the above centers plan to provide SPASE descriptions of their holdings.

System Model for Data Access



- Information flows from repositories to applications through access points and gateways in the form of SPASE descriptions as indicated by the “S”
- Registries contain inventories of resources available at one or more repositories.
- “Virtual Observatories” may be Gateways or Access Points.
- Access points may map internal metadata into SPASE descriptions.

Example Adopter: Virtual Space Physics Observatory Observatory Product Finder






GODDARD SPACE FLIGHT CENTER
Space Physics Data Facility

+ Goddard Home
+ Visit NASA.gov

SEARCH NASA

 + 60



SPASE
inside

Virtual Space Physics Observatory

[- PRODUCT FINDER](#)
 [+ ABOUT](#)
 [+ HELP](#)
 [+ ACCESSING VSPO](#)
 [+ SERVICES](#)
 [+ FEEDBACK](#)

<p>Text Restriction</p> <input style="width: 90%;" type="text"/> <input type="button" value="Add"/>	<p>Current Product Restrictions</p> <p>No restrictions are currently set.</p>																														
<p>Time Span Restriction ⓘ</p> <p>YYYY-MM-dd or YYYY-DDD</p> <p>from: <input style="width: 60%;" type="text"/></p> <p>to: <input style="width: 60%;" type="text"/> <input type="button" value="Add"/></p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;">#</th> <th style="width: 45%;">Product Name</th> <th style="width: 50%;">Access Links</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>ACE 27-day Survey Plots</td> <td>· Polar-Wind-Geotail 'gif-walk' site</td> </tr> <tr> <td>2</td> <td>ACE CRIS 1-hr Z=3-28 flux data</td> <td>· ACE Science Center · ACE/SIS L2 data in HDF via ftp</td> </tr> <tr> <td>3</td> <td>ACE Daily Survey Plots</td> <td>· Polar-Wind-Geotail 'gif-walk' site</td> </tr> <tr> <td>4</td> <td>ACE EPAM 1-hour particle flux data</td> <td>· ACE Science Center (ASC) · in HDF via ftp from ASC · CDAWeb <input type="button" value="get data"/> · in CDF via ftp from CDAWeb · in ASCII via ftp from NSSDC</td> </tr> <tr> <td>5</td> <td>ACE EPAM 5-min particle flux data</td> <td>· ACE Science Center (ASC) · in HDF via ftp from ASC</td> </tr> <tr> <td>6</td> <td>ACE GSE 12-min Position Data</td> <td>· in CDF via ftp from CDAWeb · Satellite Situation Center · CDAWeb <input type="button" value="get data"/></td> </tr> <tr> <td>7</td> <td>ACE MAG 1-hr Key Parameter (recent) data</td> <td>· in CDF via ftp from CDAWeb · CDAWeb</td> </tr> <tr> <td>8</td> <td>ACE MAG 1-hr magnetic field data</td> <td>· ACE Science Center (ASC) · in HDF via ftp from ASC · CDAWeb <input type="button" value="get data"/> · in CDF via ftp from CDAWeb · in ASCII via ftp from NSSDC</td> </tr> <tr> <td>9</td> <td>ACE MAG 16-s Key Parameter (recent) data</td> <td>· in CDF via ftp from CDAWeb · CDAWeb <input type="button" value="get data"/></td> </tr> </tbody> </table>	#	Product Name	Access Links	1	ACE 27-day Survey Plots	· Polar-Wind-Geotail 'gif-walk' site	2	ACE CRIS 1-hr Z=3-28 flux data	· ACE Science Center · ACE/SIS L2 data in HDF via ftp	3	ACE Daily Survey Plots	· Polar-Wind-Geotail 'gif-walk' site	4	ACE EPAM 1-hour particle flux data	· ACE Science Center (ASC) · in HDF via ftp from ASC · CDAWeb <input type="button" value="get data"/> · in CDF via ftp from CDAWeb · in ASCII via ftp from NSSDC	5	ACE EPAM 5-min particle flux data	· ACE Science Center (ASC) · in HDF via ftp from ASC	6	ACE GSE 12-min Position Data	· in CDF via ftp from CDAWeb · Satellite Situation Center · CDAWeb <input type="button" value="get data"/>	7	ACE MAG 1-hr Key Parameter (recent) data	· in CDF via ftp from CDAWeb · CDAWeb	8	ACE MAG 1-hr magnetic field data	· ACE Science Center (ASC) · in HDF via ftp from ASC · CDAWeb <input type="button" value="get data"/> · in CDF via ftp from CDAWeb · in ASCII via ftp from NSSDC	9	ACE MAG 16-s Key Parameter (recent) data	· in CDF via ftp from CDAWeb · CDAWeb <input type="button" value="get data"/>
#	Product Name	Access Links																													
1	ACE 27-day Survey Plots	· Polar-Wind-Geotail 'gif-walk' site																													
2	ACE CRIS 1-hr Z=3-28 flux data	· ACE Science Center · ACE/SIS L2 data in HDF via ftp																													
3	ACE Daily Survey Plots	· Polar-Wind-Geotail 'gif-walk' site																													
4	ACE EPAM 1-hour particle flux data	· ACE Science Center (ASC) · in HDF via ftp from ASC · CDAWeb <input type="button" value="get data"/> · in CDF via ftp from CDAWeb · in ASCII via ftp from NSSDC																													
5	ACE EPAM 5-min particle flux data	· ACE Science Center (ASC) · in HDF via ftp from ASC																													
6	ACE GSE 12-min Position Data	· in CDF via ftp from CDAWeb · Satellite Situation Center · CDAWeb <input type="button" value="get data"/>																													
7	ACE MAG 1-hr Key Parameter (recent) data	· in CDF via ftp from CDAWeb · CDAWeb																													
8	ACE MAG 1-hr magnetic field data	· ACE Science Center (ASC) · in HDF via ftp from ASC · CDAWeb <input type="button" value="get data"/> · in CDF via ftp from CDAWeb · in ASCII via ftp from NSSDC																													
9	ACE MAG 16-s Key Parameter (recent) data	· in CDF via ftp from CDAWeb · CDAWeb <input type="button" value="get data"/>																													
<p>Element Restriction ⓘ</p> <p>Resource type ⓘ</p> <p>Measurement type ⓘ</p> <p>Project ⓘ</p> <p>Observatory name ⓘ</p> <p>Instrument ⓘ</p> <p>Observed region ⓘ</p> <p>Instrument region ⓘ</p> <p>Spectral range ⓘ</p> <p>Timespan ⓘ</p> <p>Cadence ⓘ</p> <p>Repository Name ⓘ</p> <p>Access rights ⓘ</p> <p>Format ⓘ</p>																															

http://vsfo.gsfc.nasa.gov/websearch/dispatcher

Application Tools



Validator

Determines compliance with SPASE data model.

XML Validate: Test a web accessible SPASE description against a selected version of the data dictionary.

Parser

Convert SPASE descriptions to internal representations.

Parser: Parse SPASE XML

Generator

Create SPASE descriptions using external sources of information.

Ruleset Description Generator: Simple scripting language and templates.

Wrapper

Convert or embed SPASE metadata in other descriptions or form.

Data Dictionary Lookup: Output SPASE data dictionary in XML.

SPASE-to-OAI mapping: Map SPASE metadata to the Open Archive Initiative (OAI) metadata using stylesheets.

All tools available from: <http://www.spase-group.org/tools/>

Example Tools

Instrument Description



```
<?xml version="1.0" encoding="UTF-8"?>
<Spase xmlns:x0="http://www.spase-group.org/data/schema/spase-1_1_0.xsd">
<Version>1.1.0</Version>
<Instrument>
  <ResourceID>CDPP:IR:WINDTNR</ResourceID>
  <ResourceHeader>
    <ResourceName>Wind Waves Thermal Noise Receiver</ResourceName>
    <ReleaseDate>2006-09-01T00:00:00</ReleaseDate>
    <Description>The Thermal Noise Receiver (TNR) on the WIND spacecraft is an instrument
of a new generation, which is flown for the first time. It utilizes a digital spectrum analysis
technique similar to a wavelet transform analysis. This permits a high time resolution,
associated to a high frequency resolution. It was essentially designed to measure the thermal
noise at a high rate. The interest of this receiver also covers other events, both in situ (plasma
waves, ion acoustic waves) and remotely detected (solar and magnetospheric radio bursts).
However, this receiver uses a compression technique which limits the study of high intensity events.
Namely, the digital dynamic range is about 45 dB, added to an AGC range of more than 60 dB, the AGC
acting as a sensitivity adjustment. (e.g. an 80 dB event will not saturate the receiver, but the computed
spectrum will only cover its upper 45 dB.)
    </Description>
    <Acknowledgement/>
    <Contact>
      <PersonID>CDPP:PR:IssautierIsabelle</PersonID>
      <Role>CoInvestigator</Role>
    </Contact>
  </ResourceHeader>
  <InstrumentType>SpectralPowerReceiver</InstrumentType>
  <InvestigationName>Thermal Noise Receiver</InvestigationName>
  <ObservatoryID>spase://observatory/example</ObservatoryID>
</Instrument>
</Spase>
```

Summary / Future Plans



- The Version 1.1.0 SPASE Data Model has been released and frozen for usage and feedback from the community
- Data are being described according to the Data Model and feedback about the utility of the Model provided
- Changes necessary to the Model will be incorporated in future versions
- Tools to aid in creating resource descriptions are being developed
- Future versions of the Model will incorporate
 - model data, software descriptions, document descriptions
 - more detailed descriptions of physical parameters
 - specifications and direct access to tools for aiding descriptions
- Results and current work can be found at:
<http://www.spase-group.org>

ABSTRACT



The SPASE Data Model: Standard Metadata for Space Science Data Description

J.R. Thieman, T. King, D. A. Roberts, J. H. King, C. Harvey, and P. Richards

The Space Physics Archive Search and Extract (SPASE) project is a NASA-funded effort to create a data model for the Space and Solar Physics Virtual Observatory data environment that can be used as a common basis for locating and retrieving data of interest for the science community. Common terminology that maps to much of the disparate metadata being used by data providers in this community enables unified searches across the archives and ready comparison of the results to determine time overlaps, data commonalities, applicability for research purposes, etc. The SPASE Data Model Version 1.1.0 is presently available and updates to it will continue on a regular basis. The Data Model has been created through more than a year's collaborative efforts by representatives of worldwide space data repositories. The model now needs to be tested by the community through description of a wide variety of data holdings and feedback about the utility of the Model in enabling these descriptions. The latest version of the Data Model can be obtained by clicking on the Link marked "Current Draft" at the following web site: <http://www.igpp.ucla.edu/spase/>. Use of this model in conjunction with present and planned Virtual Observatories will be described.