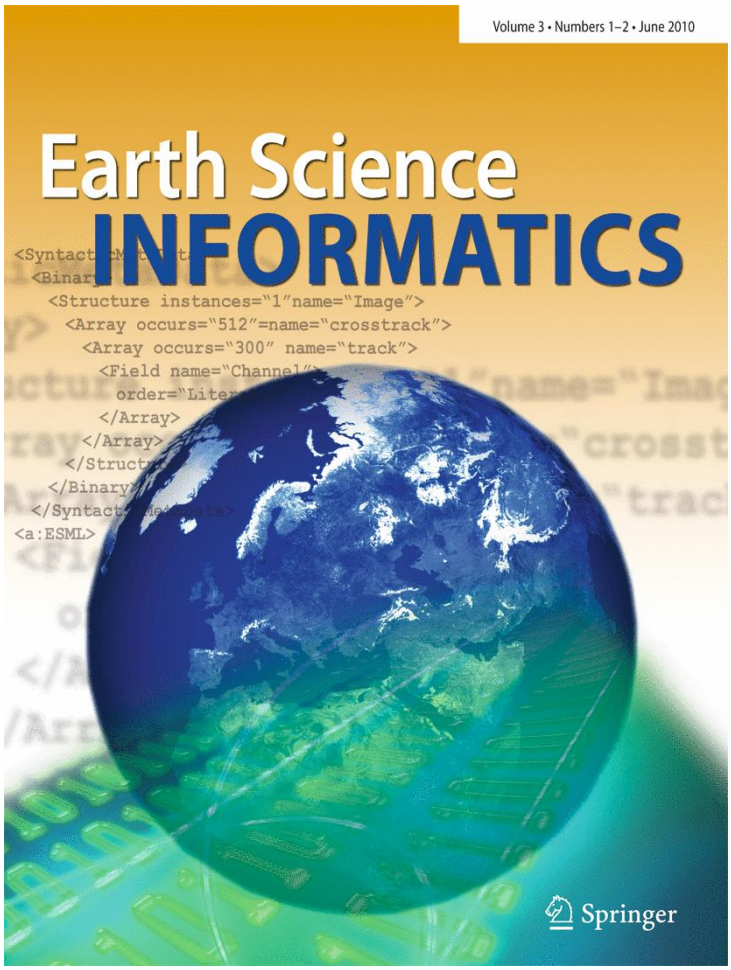
ISSN 1865-0473, Volume 3, Combined 1-2



This article was published in the above mentioned Springer issue.
The material, including all portions thereof, is protected by copyright;
all rights are held exclusively by Springer Science + Business Media.
The material is for personal use only;
commercial use is not permitted.
Unauthorized reproduction, transfer and/or use
may be a violation of criminal as well as civil law.

Earth Sci Inform (2010) 3:67–73
DOI 10.1007/s12145-010-0053-4

Author's personal copy

RESEARCH ARTICLE

SPASE 2.0: a standard data model for space physics

Todd King · James Thieman · D. Aaron Roberts

Received: 29 September 2009 / Accepted: 15 April 2010 / Published online: 5 May 2010
© Springer-Verlag 2010

Abstract SPASE—for Space Physics Archive Search and Extract—is a group with a charter to promote collaboration and sharing of data for the Space Plasma Physics community. A major activity is the definition of the SPASE Data Model which defines the metadata necessary to describe resources in the broader heliophysics data environment. The SPASE Data Model is primarily a controlled vocabulary with hierarchical relationships and with the ability to form associations between described resources. It is the result of many years of effort by an international collaboration (see http://www.spase-group.org) to unify and improve on existing Space and Solar Physics data models. The genesis of the SPASE group can be traced to 1998 when a small group of individuals saw a need for a data model. Today SPASE has a large international participation from many of the major space research organizations. The design of the data model is based on a set of principles derived from evaluation of the existing heliophysics data environment. The development guidelines for the data model are consistent with ISO-2788 (expanded in ANSI/NISO Z39.19) and the administration for the data model is comparable to that described in the ISO standards ISO-11179 and ISO-20943. Since the release of version 1.0 of the data model in 2005, the model has undergone a series of evolutions. SPASE released version 2.0 of its data model in April 2009. This version presents a significant change from the previous release. It includes the capability to describe a wider range of data products and to describe 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 describing data products. Version 2.0 of the SPASE Data Model provides a solid foundation for continued integration of worldwide research activities and the open sharing of data.

Keywords SPASE · Data model · Ontology · Vocabulary · Heliophysics · Space physics · Standard

Communicated by: Thomas Narock

T. King (✉)
Institute of Geophysics and Planetary Physics,
University of California, Los Angeles,
Los Angeles, CA, USA
e-mail: tking@igpp.ucla.edu

J. Thieman
NASA Goddard Space Flight Center, Code 690.1,
Greenbelt, MD, USA
e-mail: james.r.thieman@nasa.gov

D. A. Roberts
NASA Goddard Space Flight Center, Code 672,
Greenbelt, MD, USA
e-mail: aaron.roberts@nasa.gov

Introduction

Space Physics Archive Search and Extract (SPASE) is the moniker of a group of representatives from a number of space physics data archives, space science virtual observatories (data systems created for the purpose of providing access to metadata and data in a particular discipline, Weigel et al. 2009) and other data providers. The goal of the SPASE group is to establish standards to enable and enhance collaboration and sharing of space physics data. The main focus of the SPASE group has been the definition of a data model for space physics data. The group also undertakes tasks to demonstrate the viability of the data model and its ability to enable interoperability among data providers, archives and data services which serve the space physics community. The goal is to make it possible for data and other resources in the data environment to be easily registered, found, accessed, and used.

The recognition that a coordinated effort for the space physics data environment was necessary was first discussed at the International Solar-Terrestrial Physics

meeting at the Rutherford Appleton Laboratory in the United Kingdom in 1998. Informal, grassroots efforts were carried on within the community for the next 7 years building the foundations of a data model and defining the needs of the space physics community (Harvey et al. 2004). Participation in the SPASE consortium has grown from the initial individuals who saw the need, to international participation including groups from the United States, Great Britain, France, Canada, Japan and South Africa (see Table 1). The diversity of viewpoints has also expanded. Experts participate from a variety of space physics domains such as Magnetospheres, Waves, Ionosphere–Thermosphere–Mesosphere, Radiation, Belts, Energetic Particles, Solar Physics, and Models and Simulations. Participation in the SPASE Consortium is open to the entire community and we anticipate the membership to change continuously into the future.

The progress from concept to specifications was relatively slow in the early years of SPASE. It was recognized that a more intense effort with funding for personnel to spend a significant fraction of their time on this effort was needed and, in 2005, a proposal to create the Space Physics Archive Search and Extract project was funded through the NASA Living With a Star program. This provided support for the U. S. participation and a more intense effort was undertaken. At this time it was determined that the main part of the SPASE effort should be devoted to a common SPASE Data Model to serve as a metadata medium of information and data exchange among the widespread and diverse data archives within the community. The first operational version of the SPASE Data Model (version 1.0) was released in November 2005. In 2005 NASA solicited proposals to establish thematic virtual observatories for the heliophysics community and the SPASE Data Model was adopted as the metadata standard to enable

Table 1 Institutions with members who are participating in the SPASE Group

Augsburg College http://www.augsburg.edu/home/physics/spacephysics.htm
Canadian Space Science Data Portal http://www.cssdp.ca/
CNES/CNRS Plasma Physics (CDPP) Data Archive http://www.cnes.fr/
European Grid of Solar Observations (EGSO) http://www.egso.org/
Inter-university Upper atmosphere Global Observation NETwork (IUGONET) http://www.iugonet.org/
Institute of Geophysics and Planetary Physics – UCLA http://www.igpp.ucla.edu
Institute of Space and Astronautical Science (ISAS/JAXA) http://www.isas.ac.jp/e/index.shtml
Jet Propulsion Laboratory http://www.jpl.nasa.gov/
John Hopkins University - Applied Physics Laboratory http://www.jhuapl.edu/
NASA/Goddard Space Flight Center http://hpde.gsfc.nasa.gov/
NOAA/National Geophysical Data Center http://www.ngdc.noaa.gov/
Planetary Data System- Plasma Physics Interactions Node http://ppi.pds.nasa.gov/
Rutherford Appleton Laboratory http://www.scitech.ac.uk/
Southwest Research Institute http://www.swri.org/

interoperability. In 2008 the initial SPASE grant ended and NASA committed to longer term support for SPASE by making it an infrastructure component of the Heliophysics Data Environment (HPDE, http://hpde.gsfc.nasa.gov). From the inception of the SPASE group to the present more than 4000 emails, over a hundred biweekly teleconferences, and at least six face-to-face meetings, the SPASE Data Model continued to be developed and the latest version is now 2.1.0 as of March, 2010.

The SPASE Data Model is now being used by a variety of groups within the United States, Europe, Canada and Japan. In the U.S. the SPASE Data Model is an infrastructure component of NASA's Heliophysics Data Environment (HPDE). The HPDE consists of nine Heliophysics Virtual Observatories (VxO) and a growing number of service providers (e.g., Autoplot (http://www.autoplot.org), Heliophysics Event List Manager (HELM), etc.). The VxO's use the SPASE Data Model to document both existing and newly acquired resources. Service providers and the VxOs use the assigned unique resource identifiers to retrieve the resources (data) and metadata. Information from the resource descriptions is harvested to enable search engines and specialty services. The SuperMAG project of the National Science Foundation also provides SPASE Data model compliant descriptions for its resources enabling cross agency sharing of data which is something that was remarkably limited in the past.

Both the Canadian Space Science Data Portal (CSSDP) and the Cluster Active Archive (CAA) in Europe are using the SPASE Data Model. The European Union's HELIO project will be able to consume SPASE Data Model compliant descriptions and has intentions to provide SPASE Data Model compliant descriptions. In Japan the Inter-university Upper atmosphere Global Observation NETwork (IUGONET) has chosen SPASE as their metadata exchange standard.

SPASE data model development

The SPASE Data Model is both a conceptualization of a domain and a reflection of the intrinsic aspects of the data environment. For the space physics domain, data originates and is curated by many different organizations, exists in many different formats, and is continuously augmented. These intrinsic aspects can be stated as a set of principles which guide the design, development and continued enhancement of the SPASE Data Model.

- **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.
- **An Online Resource has a Universal Resource Locator (URL)**

  If a resource is on-line it can be accessed and retrieved using a Universal Resource Locator (URL).
- **The data environment is continuously evolving.**

  New resources are actively generated either as part of an on-going experiment or as a result of analysis and assessment.

Along with the design principles a basic tenet was to allow the data model to evolve based on necessity. The earliest version of the data model (released in 2005) was intentionally rudimentary. It contained basic definitions for Person, Observatory, Instrument, NumericalData, DisplayData and Catalog classes. Extensions or expansions of the basic model occurred as existing resources were described using SPASE; occasionally a resource could not be adequately described with the existing data model. The issues were raised and discussed by the SPASE group and suggestions for improvements were formulated. When a consensus was reached the changes were made to the data model. This has resulted in an organic, yet controlled, evolution to the SPASE data model.

The specification of the SPASE Data Model uses terminology which is slightly unconventional, but is easily mapped to current standardized nomenclature. This is in part due to the timeframe of the origins of SPASE (1998) and the decision to adopt terminology commonly understood by the SPASE community. The SPASE community is lead by scientists who are assisted by computer scientists, data engineers, and data developers working within the space physics domain. One nomenclature example is that the SPASE consortium uses the terms "Resource", "Element", and "Container" which would correspond respectively to the terms "Class", "Attribute", and "Component Class" within the discipline-independent Unified Modeling Language (UML). The SPASE consortium also defined a set of development procedures and policies which have been found to align well with several international standards for metadata.

The SPASE Data Model is primarily a controlled vocabulary with hierarchical relationships and the ability to form associations between described resources. The selection of element names (dictionary terms) in the controlled vocabulary follows the guidelines outlined in (ISO-2788 1986) and expanded in (ANSI/NISO Z39.19 2005). Element names are also consistent with naming and identification principles described in (ISO-11179-5 2005). That is, each element name in the vocabulary has one and only one meaning; element names are always in singular form; compound element names are used as necessary to convey the semantics of the element; and lists (enumerations) are used to clearly differentiate aspects of a conceptual set. For example, the element to capture how readily a resource can be accessed has the name

"Availability". Some elements have multi-word (compound) names to capture the full semantics of the element. For example, the class of "range" elements which includes "Azimuthal Angle Range", "Polar Angle Range", "Energy Range", "Wavelength Range", and "Frequency Range".

Terms are organized into whole-part relationships (classes). The SPASE Data Model employs both hierarchical relationships (taxonomy) of classes and associative relationships. Some classes have polyhierarchical relationships in that they can be included in more than one other class. Every term in the SPASE vocabulary (dictionary) has a single definition which is compliant with the Data Element Formulation Rules detailed in (ISO-11179-4 2004)

The administration of the SPASE Data Model adheres to the principles of ISO-11179 (Metadata Registry) standard, although SPASE does not formally adopt all aspects of this standard. Each element in the SPASE Data Model is a Data Element Concept with fully specified representation (value domain and data type) and the terms are organized into object classes (ISO-11179-1 2004). Each item (element, container, enumeration) in the SPASE data model is an Administered Item (ISO-11179-3 2004) and is assigned a version number which is synchronized with the classification scheme of a data model release.

The procedures for the administration of the data model follow closely the specification found in ISO-20943 "Procedures for achieving metadata registry (MDR) content consistency" (ISO-20943-1 2003). When new elements are suggested an evaluation is performed to determine if elements with a comparable concept already exists and if an existing international, national or organizational standard exists. For example, SPASE adopted the International Organization of Standards (ISO) standard ISO-8601 for date, time and duration specification (ISO-8601 2004). SPASE also adopted the specification for a Uniform Resource Identifier (URI) (Berners-Lee et al. 2005) for unique identifiers that are assigned to each described resource. A URI that has the form

scheme://authority/path

where "scheme" is "spase" for those resources administered through the SPASE framework, "authority" is the unique identifier for the naming authority within the data environment and "path" is the unique local identifier of the resource within the context of the "authority". Since SPASE manages all aspects of its data model ISO-11179 administered item attributes like "Submitting Organization" and "Stewardship Contact" are implied element attributes and not maintained in the SPASE dictionary.

The Data Model specification is implementation neutral and does not require any particular implement. However, the reference implementation chosen by the SPASE consortium uses XML (eXtensible Markup Language). The SPASE group generates an XML schema for each version of the data model and a corresponding XML Metadata Interchange (XMI) documents, which can be used in Unified Model Language (UML) tools. There are also XML style sheets for converting the metadata to HTML or OAI (Open Archives Initiative). Tools for exploring the data model are also available. This includes a dictionary search capability, a linear tree view and a tree based explorer (available from http://www.spase-group.org/).

SPASE 2.0

Version 2.0 of the SPASE Data Model is a significant change from the previous release. It includes the capability to describe a wider range of data products and infrastructure components than was possible with Version 1.0. The top-level class hierarchy of the SPASE Data Model is shown in Fig. 1. A SPASE description can consist of any number of primary classes (conceptually referred to as a "resource"). The primary classes for data resources are Annotation, Catalog, Document, DisplayData (images), Granule, and NumericalData. A data resource describes data product sets which are composed of one or more products. The primary classes for entities are Instrument, Observatory and Person. Entity Resources describe the generators or sources of data. The primary classes for infrastructure components are Registry, Repository and Service. Infrastructure Resources describe system components that are part of the exchange and use of data. All classes, with the exception of Granule and Person, share attributes that are encapsulated as a Resource Header class.

Each of the data classes can include any number of Parameter classes. The Parameter class (see Fig. 2) describes a set of sampled values which comprise a resource. A resource may have any number of parameters. Each parameter is characterized as a particular observed phenomenon or measured quantity. A Parameter may be a Field, Particle, Wave, Support value or Mixed (calculated) value. All Parameters can have a CoordinateSystem, RenderingHints and an internal Structure. Not shown in either class diagram are other supplemental classes that are associated with these classes and the attributes of each class. The full details of the SPASE Data Model are available at http://www.spase-group.org/data/doc/spase-2_0_0.pdf.

Enhancements which lead to Version 2.0 include the addition of terms to support both active and passive wave related data, an important distinction for wave studies in space physics. The model for the components of multi-dimension data was expanded from an exclusively Cartesian model (classic X,Y,Z) to support angular components used in cylindrical and spherical coordinate systems. An Annotation resource was added to allow additional information to be associated with existing resources. This

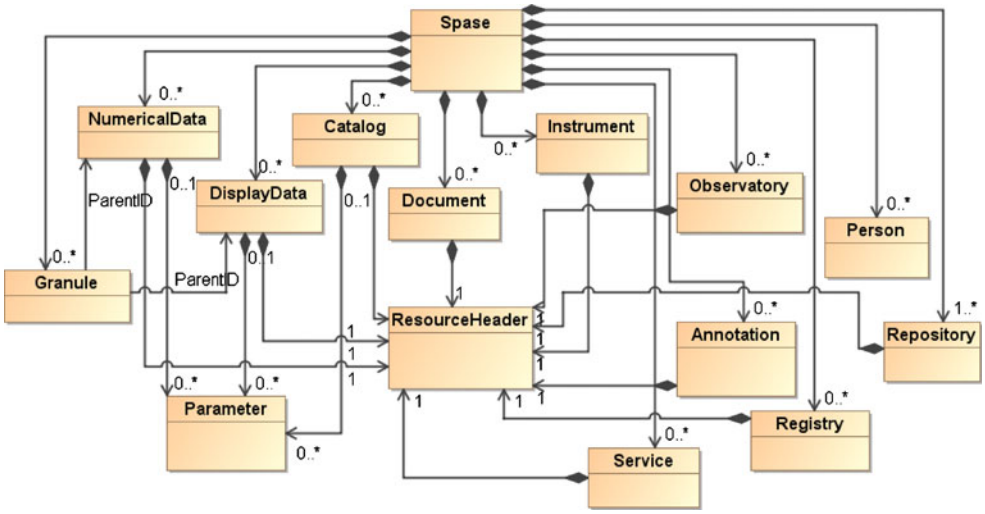


Fig. 1 A top-level class diagram of the SPASE data model. The notation near an arrow head indicates the cardinality (how many occurrences) are allowed for the class. A value of "1" indicates that it is required, "0..*" indicated any number are allowed. The closed diamond indicates a composition relationship where the referenced classes are destroyed along with the "whole." A data model compliant description can consist of any number of primary classes (conceptually referred to as a "resource"). Data related classes share attributes encapsulated as a Resource Header class and may have multiple associated Granule classes. The contents of data resources are described with one or more Parameter classes. Not shown are attributes and some supplemental classes

information can be typed to describe an event, anomaly or feature discovered in a resource. The Granule resource, which is used to describe individual files that are part of a conceptual resource (for example a NumericalData resource), was enhanced to support the tight coupling of related sources (files). This includes thumbnail and browse images, ancillary metadata, and layout information stored as separate files. Additional improvements were made to enhance the capability to describe resource associations. One resource can now be associated with another with relationships of Child, Event Of, Derived From, Observed By, Part Of, Revision Of and Other.

SPASE usage

As an example of the SPASE Data Model in the XML reference implementation, below is a description of an instrument on the Geostationary Operational Environmental Satellite (GOES) 10 spacecraft. As with every resource a unique identifier is assigned to the resource and is specified by the ResourceID tags. Two other resources are referenced by this document. A Person resource in the Contact segment and the Observatory (spacecraft), which is the host for the instrument, in the ObservatoryID tags. The description has been truncated to keep the example brief.

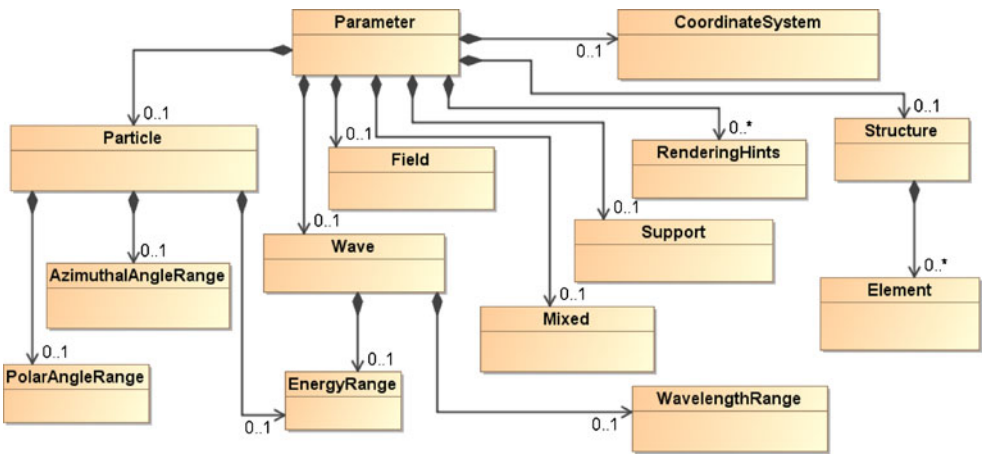


Fig. 2 The Parameter class diagram. A parameter class describes the sets of sampled values which comprise a resource. Not shown are attributes and some supplemental classes

```
<?xml version="1.0" ?>
  <Spase xmlns="http://www.spase-group.org/data/schema">
  <Version>2.0.2</Version>
  <Instrument>
    <ResourceID>spase://SMWG/Instrument/GOES10/MAG</ResourceID>
    <ResourceHeader>
      <ResourceName>GOES 10, Triaxial Fluxgate
Magnetometer</ResourceName>
      <ReleaseDate>2009-11-21T00:46:11Z</ReleaseDate>
      <Description>The following brief instrument description,
written after the
 launch of GOES 8 and 9, but that applies to GOES 8-12, is
reproduced, in part,
from:

Singer, H.J., L. Matheson, R. Grubb, A. Newman and S.D. Bouwer,
Monitoring Space Weather with the GOES Magnetometers. SPIE
Conference Proceedings, Vol. 2812, p. 299-308, GOES-8 and Beyond,
Edward R. Washwell, ed., 1996

The magnetometers on the three-axis stabilized GOES I-M (called
GOES 8-12 after launch) series of satellites are three-axis
fluxgate magnetometers manufactured
… description abbreviated …
</Description>
      <Contact>
        <PersonID>spase://SMWG/Person/Howard.J.Singer</PersonID>
        <Role>PrincipalInvestigator</Role>
      </Contact>
    </ResourceHeader>
    <InstrumentType>Magnetometer</InstrumentType>
    <InvestigationName>GOES 10, Triaxial Fluxgate Magnetometer on
GOES 10</InvestigationName>
    <ObservatoryID>spase://SMWG/Observatory/GOES10</ObservatoryID>
  </Instrument>
</Spase>
```

The future

Modification and improvement of the Data Model remain the prime tasks for the SPASE consortium as users report on their experiences in the use of the model. The SPASE group is also working to improve support and services including better metadata editors, improved stylesheets, standardization of registry protocols and tools to aid the migration of metadata to new versions of the data model and to other data model standards, for example, the Planetary Data System (http://pds.nasa.gov). Other services will be developed as the needs are identified. In addition, the documentation for the SPASE Data Model is undergoing a process of evolution to improve the information available to both new and experienced users. This includes tutorials, guideline documents, and references to examples and helpful auxiliary information.

Development is also proceeding on a number of services to be offered to the community (see http://www.spase-group.org/tools). One central tool is the SMWG (SPASE Metadata Working Group) Resource Registry which serves as a ready source of information about the spacecraft, instruments, personnel, and other entities that can be used in data related descriptions. There will also be interfaces to reports that could be useful for data descriptions as well. When data have been found and extracted it is always useful to have data visualization capabilities, so services such as Autoplot for graphic display have been updated to

understand SPASE Resource Identifiers for accessing data resources. A search capability across the space physics domain is also a valuable aid to those wishing to describe data. A centralized search service for NASA resources is located at the Virtual Space Physics Observatory (http://vspo.gsfc.nasa.gov). To carry out searches across agencies and between thematic virtual observatories two Application Programmer Interfaces are presently in consideration, the SPASE Query Language or SPASE-QL (Narock and King 2008) and one which uses Representational State Transfer (REST) that is part of the SPASE registry service (http://www.spase-group.org/tools/collection/).

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 full and concise documentation utilizing SPASE technologies can be quick and (nearly) effortless. The SPASE group is working towards this goal. For more details, please visit the SPASE website found at: http://www.spase-group.org.

The SPASE community welcomes additional participation from those who have data in this domain or just an interest in being involved in the data community. If you have resources such as data, event lists, or anything that might be useful to the space physics data environment contact one of the authors of this paper or representatives of the participating virtual observatories or a member of one of the institutions listed in Table 1 to get started. If you have a lot of resources to share, consider establishing a personal virtual observatory within the domain and having a SPASE registry in order to join the federated environment.

Acknowledgments This work was supported in part by the National Aeronautics and Space Administration under Grant No. NNX09AF15G.

References

ANSI/NISO Z39.19 (2005) Guidelines for the construction, format, and management of monolingual controlled vocabularies

Berners-Lee T, Fielding R, Masinter L (2005) Uniform Resource Identifier (URI): generic syntax, internet engineering task force, RFC 3986, http://tools.ietf.org/html/rfc3986

Harvey CC, Thieman JR, King T, Roberts DA (2004) SPASE—space physics archive search and extract, In Proceedings PV-2004 Ensuring the Long Term Preservation and Adding Value to Scientific and Technical Data, Frascati, Italy, ESA/ESRIN WPP-232

Narock TW, King TA (2008) Developing a SPASE query language, Earth Science Informatics, Issue: 1, 2008 doi:10.1007/s12145-008-0007-2

ISO-2788 (1986) Guidelines for the establishment and development of monolingual thesari

ISO-8601 (2004) Data elements and interchange formats—information interchange—representation of dates and times

ISO/IEC 11179-1:2004(E) (2004) Information technology—metadata registries (MDR), Part1: Framework

ISO/IEC 11179-3:2003(E) (2003) Information technology—metadata registries (MDR), Part3: Registry metamodel and basic attributes

ISO/IEC 11179-4:2004(E) (2004) Information technology—metadata registries (MDR), Part4: formulation of data definitions

ISO/IEC 11179-5:2005(E) (2005) Information technology—metadata registries (MDR), Part5: naming and identification principles

ISO/IEC 20943-1 (2003) Information technology—procedures for achieving metadata registry (MDR) content consistency

Weigel RS, Baker DN, Roberts DA, King T (2009) Using virtual observatories for heliophysics research. EOS Transactions 90:441–442