Roberts D. Aaron (Orcid ID: 0000-0001-6565-2921)
King Todd (Orcid ID: 0000-0002-0523-627X)
Fung Shing F. (Orcid ID: 0000-0002-1216-5553)
Cecconi Baptiste (Orcid ID: 0000-0001-7915-5571)

The SPASE Data Model: A Metadata Standard for Registering, Finding, Accessing, and Using Heliophysics Data Obtained from Observations and Modeling

D. Aaron Roberts[1], James Thieman[2], Vincent Génot[3], Todd King[4], Michel Gangloff[3], Chris Perry[5], Chiu Wiegand[6], Darren De Zeeuw[7], Shing F. Fung[8], Baptiste Cecconi[9], and Sébastien Hess[10]

[1] Heliophysics Science Division, Code 672, NASA GSFC, Greenbelt, MD 20771, USA.; aaron.roberts@nasa.gov. [2] NASA Goddard Space Flight Center, Code 690.1, Greenbelt, USA. [3] IRAP, Université de Toulouse, CNRS, UPS, CNES, Toulouse, France. [4] Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA, USA. [5] RAL Space, STFC Rutherford Appleton Laboratory, Harwell Oxford, OX11 0QX, United Kingdom. [6] Heliophysics Science Division, Code 674, NASA GSFC, Greenbelt, MD 20771, USA. [7] Climate & Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. [8] Heliophysics Science Division, Code 675, NASA GSFC, Greenbelt, MD 20771, USA [9] LESIA, Observatoire de Paris, 5 place Jules Janssen, 92195, Meudon Cedex, France [10] French Aerospace Lab., ONERA, Toulouse, France.

Corresponding author: D. Aaron Roberts (aaron.roberts@nasa.gov)

Key Points:

- The Heliophysics/Space Weather community has developed a general means to register, discover, access, and use datasets and other products.
- The key for ease-of-use is the adoption of standards for data formats, metadata, and access methods.
- The “Space Physics Archive Search and Extract” Data Model provides a stable and general standard for the requisite metadata descriptions.

Abstract

The Space Physics Archive Search and Extract Consortium has developed and implemented the "SPASE Data Model" that provides a common language for registering a wide range of Heliophysics (HP) data and other products. The Data Model enables discovery and access tools such that any researcher can obtain data easily, thereby facilitating research, including on space weather. The Data Model includes descriptions of Simulation Models and Numerical Output, pioneered by the Integrated Medium for Planetary Exploration (IMPEx) group in Europe, and subsequently adopted by the Community Coordinated Modeling Center (CCMC). The SPASE group intends to register all relevant Heliophysics data resources, including space-, ground-, and model-based. Substantial progress has been made, especially for space-based observational data and associated observatories, instruments, and display data. Legacy product registrations and access go back more than 50 years. Real-time data will be included. The NASA portion of the SPASE group has funding that assures continuity in the upkeep of the Data Model and aids with adding new products. Tools are being developed for making and editing data descriptions. Digital Object Identifiers (DOIs) for Data Products can now be included in the descriptions. The data access that SPASE facilitates is becoming more uniform and work is progressing on Web Service access via a standard Application Programming Interface. The SPASE Data Model is stable; changes over the past nine years were additions of terms and capabilities that are backward compatible. This paper provides a summary of the history, structure, use, and future of the SPASE Data Model.

1 Motivation and Overview

Over three decades ago, what NASA now calls the Heliophysics (HP) community already knew what was needed to have orderly general access to the data produced by its many missions (National Research Council, Committee on Data Management and Computation, 1982—the "CODMAC Report"; also, Bogart, et al., 1998). In current terms, the data system would need to be standards-based, with scientist input at all levels, allowing users to easily find, access, and use data from all Heliophysics systems and models. The system would have the ability to bring old data to life and to preserve raw and processed data from old and new missions. The standards for formats and access would facilitate producing general browsing and analysis tools. It took many years, the advancement of computer technology, and the gradual buy-in by the HP community to make substantial progress on achieving these goals, but we now have a NASA Heliophysics Scientific Data Management Policy (NASA HPDD, 2016) that provides a basis for progress. One major step has been the establishment of a standard for metadata to describe HP data and its sources—the SPASE Data Model—that provides a route to discovery and access for the various resources in Heliophysics including data from observatories and models, catalogues (including event lists), and overview "display" data in the form of prepared graphs. Our work is consistent with the widely adopted "FAIR" principles (Findable, Accessible, Interoperable, and Re-useable: see https://www.force11.org/group/fairgroup/fairprinciples and: https://www.nature.com/articles/sdata201618).

The general structure of the Heliophysics Data Environment, while grounded in the ideas of the "CODMAC" report, is now based on the idea of a "Virtual Observatory" (VO) which extends the earlier ideas given what is possible with new technology. The astronomical community realized nearly two decades ago that taking advantage of existing observations stored throughout the community drastically changed the amount of data available to a researcher. Initially termed a "Digital Sky," this became the National Virtual Observatory (see https://hea-www.cfa.harvard.edu/USVOA/) in the US, and the International Virtual Observatory Alliance (IVOA; http://ivoa.net) worldwide. As the earlier efforts had foreseen, to be successful, standards were required primarily for: (1) data formats, (2) metadata descriptions, and (3) access methods (Application Programming Interfaces, or APIs). A primary goal of the IVOA is to make all observations, in whatever wavelength, appear the same as all others in terms of discovery, access, and, to the extent possible, use. Uniformity of the data format, metadata, and APIs makes it possible for client software (in IDL, MatLab, Python, or other languages or tools) to load data arrays using just dataset and variable IDs as if from a local disk. Given the infrastructure facilitated by standards, Web portals (also a possible API use, but with direct user interaction) allow one easy route to data, and these can provide everything from quick-looks that enhance research efficiency to the ability to make publication-ready plots and images.

Heliophysics is not yet quite as unified as Astrophysics, but what started as a wide variety of idiosyncratic *formats* has become FITS for solar images, CDF for space physics data, and NetCDF for some major datasets closer to the Earth in the Ionosphere-Mesosphere-Thermosphere realm. (ASCII is decreasingly used as an archival format.) *Access* has been improved with the Virtual Solar Observatory API providing nearly universal solar access, and a few APIs (particularly "CDAS" for the CDAWeb archive; https://cdaweb.gsfc.nasa.gov/WebServices/) providing access to much space physics and some ground-based data. The "Heliophysics API" ("HAPI"; https://github.com/hapi-server/) has now been defined and is being implemented with the intention of providing a single route to all HP time series data with easily adopted methods for servers and clients. Most central to this paper, the *metadata* for HP has been defined for many years now using the SPASE Data Model (hereafter simply "SPASE"). Two decades of effort has gone into defining and refining SPASE, with input from all HP disciplines and, from the start, with international collaboration. One significant area of effort is relating the European use of IVOA standards to SPASE and HAPI. Ultimately the users should not see any of the details; they will simply request and use data via a common route. For example, AMDA (a space plasmas analysis tool, http://amda.cdpp.eu) and 3DView (a spacecraft trajectory, in-situ data, and model visualization tool, http://3dview.cdpp.eu) have been using metadata compliant with SPASE for several years to build their hierarchy (or tree) of data for observations as well as simulations (see for instance Génot et al., 2017).

The next section provides historical background on SPASE, but those only interested in the content of the data model can skip to subsequent sections that deal with the structure and uses of SPASE for both observations and models. The NASA context of this work is given in NASA HPDD (2016).

2 A Brief history of SPASE

Earlier presentations of SPASE and related tools are given by Harvey et al. (2008), King et al. (2010), and Thieman et al. (2010); these give some idea of the international origins and development of SPASE. The SPASE effort has its roots in a call for action at a data handling session of the International Solar Terrestrial Probes (ISTP) workshop held at the Rutherford Appleton Labs (RAL) in 1998, when on September 26 a resolution was passed calling on the "larger data centers" to "do something" to make data more accessible. This was to be a continuation of the work mentioned above that started with CODMAC and continued with other meetings. RAL, CDPP (Centre de Données de la Physique des Plasmas, http://www.cdpp.eu), NSSDC (National Space Science Data Center), and SWRI (Southwest Research Institute) took up the challenge. Early in 2001 a breadboard interoperability test bed was implemented between NSSDC and CDPP, and later that year, NSSDC, SWRI, RAL and CDPP submitted a proposal to NASA for "A Space Physics Archive Search Engine." Although this 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.

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 by the Living With a Star (LWS) program, which helped to accelerate the effort. Finally, in November 2005, the SPASE group released version 1.0 of its "ontology" (Data Model). Further product descriptions led to the need for updated versions. In 2006, NASA established thematic "VxOs" (Virtual Observatories with "x" being "Magnetosphere," as in VMO, "ITM," being "Ionosphere, Thermosphere, Mesosphere" as in VITMO, etc.) that helped to bring focused efforts to SPASE from each of the subdisciplines. After a period of use in NASA's VxOs, the model was streamlined and enhanced to support a wider range of resources, and the first and only (so far) non-backward compatible version (2.0) was released in April 2009. Since then, the major change was to add the Simulation Extension Product Types (e.g., "SimulationRun"; see below) in May 2014. The modifications since then have all been in the "base" SPASE model, so the Simulation Extensions remain at Version 1.0.0 (see Sec. 6).

The use of SPASE to provide descriptions and tools was slow at first, but now a large fraction of HP data has been described and registered in these terms, and all current NASA missions are described by SPASE, as required by the HP Data Policy (NASA HPDD 2016). Starting in 2017, NASA has provided long-term funding, as part of the Heliophysics Data Environment (HPDE) infrastructure, to efforts that will maintain and augment both resource descriptions and the Data Model itself (see http://spase-group.org and https://hpde.gsfc.nasa.gov). Throughout this process, links have been maintained to non-NASA agencies and groups, in the US and abroad.

The rest of this paper describes the SPASE Data Model, as an introduction to make it more easily useable. Inevitably, SPASE has grown increasingly complex, but its foundation has always been to provide basic information about HP resources in ways that are simple for users to understand. The detailed specification of the current SPASE Data Model at the time of this writing is given by The SPASE Consortium (2018). This includes a complete hierarchy of the model, many example descriptions, and a dictionary of terms. The Simulation Extensions are given by The SPASE Consortium (2014).

3 SPASE Resources

Each SPASE document—a basic entry in a SPASE Registry—is a description of a "Resource," the latter being the generic term for any entry in the system. Each Resource is assigned a unique Resource ID that is used as a reference, both externally and by other resources. Figure 1 shows, in the left column, the Resource Types colored to show the main categories. Each Resource Type consists of a set of attributes that characterize the resource. Each Type also has its own SPASE descriptor format, typically encoded in XML. The format for each Product Type is given in an "XML Schema" that can be used to check a description for its conformity to SPASE requirements (see the online tool for validating SPASE syntax at http://spase-group.org/tools/validate/). Resource Types can be divided into three categories, color-coded in the Figure.

Data Resources (blue) describe one or more data products. A "data product" is a set of data that is uniformly processed and formatted, from one or more instruments, typically spanning the full duration of the observations of the relevant instruments. A data product may consist of a collection of files of successive time spans but may be high-level entities such as event catalogs. Data products can be images (Display Data), sample or observation values (Numerical Data), or event lists (Catalog). Included in the Data Resource category are the resources that describe individual files (Granules) which are part of data product sets, and assessments of a resource (Annotations). The complete list of observational Data Resources is: Numerical Data, Display Data, Catalog, Granule, and Annotation. Numerical Data Resources are typical numerical variables or quantitative images as a function of time, and Granules are the individual files or other basic division that contain these. Display Data Resources are jpeg, png, or other graphical or image representations of the Numerical Data, often used for browsing but in some instances suitable for publication. Catalogs are usually event lists that give, say, times when solar Coronal Mass Ejections occurred or times when high-resolution ("burst") data are available from one or more observatories.

Origination Resources (green) describe the generators or sources of data. People (Person Resources) are the authors/creators of all the Products, and Instruments on or in Observatories are the sources of Data. A Data Resource refers to one or more Origination Resources.

Infrastructure Resources (red) describe system components that are part of the access, transfer, and use of data. Repositories are the containers of Data Products; Services act on data or metadata; and Registries are containers of SPASE descriptions that keep track of all Products. Documents are also placed in this class as typically being supportive of the Data and other Resources. They are less formally organized than other SPASE Resources since the content of relevant papers, etc. can take many forms. Note that SPASE does not intend to register documents that are otherwise accounted for, such as journal articles, but rather things such as the NASA Data Policy or formal SPASE Data Model descriptions.

Simulation Resources are analogous to the Resources above that are based on observations. The output from simulations (Numerical Output and the related Display Output) is essentially the same as observational data, and thus these Resource Types are in blue in Fig. 1. However, instead of being generated by an Instrument of an Observatory the outputs are the result of Simulation Runs of a Simulation Model. Thus, the latter two Resource Types are included as Origination Resources in Figure 1. See below for further discussion of simulations.

4 Core attributes and their hierarchy

With the exception of Granule and Person, which are very simple in structure, every resource has a common set of core attributes. The core attributes provide textual descriptions of the resource and the capability to reference external sources of information (Information URL). It also describes the context of the resource in the larger data environment. This context consists of associations with other resources (Association) and with previous versions (Prior ID). These attributes are grouped in a Resource Header (see Figure 1) and consists of: Resource Name, Alternate Name, Digital Object Identifier (DOI; see below Sec. 7.7), Publication Information to go with the DOI, Funding (the source of support), Release Date, Expiration Date, Description, Acknowledgement, Contact, Information URL, Association, and Prior ID.

Data Resource descriptions also include overviews that tell the user other essential general information, such as what Repository, Instrument, and Observatory the Resource comes from, how to access it, what its measurement type is, the time span and cadence of the observations/output, what spatial and spectral region it applies to, and who provided it. Thus, the basic information in SPASE provides users with the ability to understand the relevance of a given Resource to his or her research, in addition to the most direct routes (via Access URLs) to obtaining data.

Figure 1 also shows some of the details of a Numerical Data Resource, including four levels of the hierarchical description. Note that People, Instruments, Repository, and (implicitly) Observatory are included by reference, the latter by the reference shown in the Instrument description hierarchy. The spase-group.org website provides various ways of seeing the full set of SPASE terms and the hierarchy. One method uses diagrams, as in Figure 2, to provide the full content of the hierarchy at each level (http://spase-group.org/data/model/spase-2_3_0/). The full hierarchy is also in the references for the main descriptions (The SPASE Consortium, 2014, 2018).

5 Parameter (data variable) descriptions

Data Products, as well as Numerical Output from simulations, contain variables of varying degrees of complexity, from simple scalars to vectors, spectrograms, images, and multidimensional quantities such as a distribution of particles by species, polar angle, azimuthal angle, and energy. A complete description of a Data Product will include specifications of all such quantities (termed "Parameters"–see Figure 1–despite other possible meanings of the word) in the product, although the Parameter descriptions are optional so that products can be described more easily to provide at least registration and search capabilities. Parameters can also include Support quantities, such as data quality flags. One unusual aspect of the description is the "Parameter Type" which is not a separate entity but just a placeholder for one of the entities on the list that is attached to it. Also of note are the purple entities that provide "keys." The use of these is in direct access APIs—see Sec. 7.4.

A current task of the SPASE group is filling in the missing Parameters in many earlier descriptions. As with the general description of the Product, the Data Model provides options for very detailed Parameter descriptions, with some elements being required if the Parameter is to be included. The elements include information concerning the data cadence, units, and coordinate system, plus the Parameter Type such as a Field, Particle, or Wave measurement. The detailed structure of the data files can be captured to allow users to be able to figure out how to use the data, no matter how complex. The richness of the SPASE Parameter schema also permits straightforward data rendering specifications for generating graphs and other figures.

6 Simulation Extension

From early on, the SPASE group realized that a complete view of HP research had to include simulations, the models they were based on, and a route to using the often-voluminous numerical output from them. Ultimately, the success of our models is that their results are consistent with observational data; this is the basic science requirement of having theories predict observations that in turn test theories. In May of 2014, it was decided to add the entries for Simulations as a SPASE "extension." The Resource Types added in this way (see Fig. 1) function in the same way as those in the "Base Model" (consisting of all the original Resources), but they are treated as a separate population so that Simulation Resources can evolve independently. Terms from the Base Model, when used in an Extension (e.g., "Resource Header," "Parameter," "Access Information") inherit all the properties of the Base Model term. Two Base Model terms, namely Granule and Particle, were overwritten (redefined) in the Extension, but all others remained the same.

The main impetus for the Simulation Extensions was the work of the Integrated Medium for Planetary Exploration (IMPEx; http://impex-fp7.oeaw.ac.at/) project, a European Union (EU)

Seventh Framework Programme sponsored project, which was subsequently endorsed by the SPASE consortium. The extension to SPASE developed by the IMPEx team was incorporated into SPASE in June of 2015. The extension for simulations was designed to be as generic as possible, but the IMPEx team focused on just a few models. New models were subsequently described in the frame of Transplanet (a simulation platform for ionosphere models http://transplanet.irap.omp.eu), using the extension with only a few additions or modifications. A French Agence Nationale de la Recherche project called TEMPETE (Temporal Evolution of Magnetized Planetary Environments during exTreme Events, P.I.: R. Modolo, LATMOS: Laboratoire Atmosphères, Milieux, Observations Spatiales; see http://www.agence-nationale-recherche.fr/Project-ANR-17-CE31-0016), aiming at simulating planetary exospheres and magnetospheres. This project started recently and will also use the Simulation Extension to describe these models.

In another modeling context, the Community Coordinated Modeling Center (CCMC) at NASA Goddard is a community resource of hundreds of models and many thousands of model runs made available to the public. The CCMC is now working to make SPASE the meta-data model for its holdings, including for the Integrated Space Weather Analysis System (iSWA; https://ccmc.gsfc.nasa.gov/iswa/) that provides near-real time datasets, as well as for the more complex model output. Each model and model version will be fully described by a separate Simulation Model resource, including authors, references, and key publications to best describe the model. The Simulation Run resource is the most complex in that it must capture all inputs to a model and how it was run with enough detail to be able to reproduce a run. Many models have a simple set of inputs, but several sophisticated models have hundreds of options and run in multiple "sessions" that change options during the run. Future integration of the Simulation Run metadata with the CCMC's Run-on-Request system will enable far more customization of options. The Numerical Output resource is very similar to its observational data counterpart (Numerical Data), although a wide variety of output with differing dimensions, cadences, and scales can make the description complicated.

The CCMC is implementing the above ideas by developing a large database called the CCMC Metadata Archive (CMA; see Fig. 3) which will act as their central repository of all the metadata to describe their data and model holdings; the current metadata reflects the variety of original model designs. When completed, this database will interact with the official SPASE registry either through APIs or through exported XML files to cooperate in the SPASE network of services. Internally to the CCMC, the CMA will become an integral part of their services, such as iSWA (a space weather real-time data viewer), Kameleon (a set of tools based on a uniform format for simulation output), runs-on-request, and metrics and validation challenges, This integration of standard metadata will allow not only more finely tuned discovery searches for data in stored runs due to, e.g., the uniformity of time formats and variable types (for example, which models output plottable O+ density, or what models can do polar outflow?), but also better visualization and analysis tools. It will enable reliable links to data APIs like HAPI to facilitate data-model comparisons. Similarly, CCMC is planning to use the SPASE metadata in the new CAMEL validation tool to help with overlaying data/model results with correct units.

7 SPASE in practice: Generation and Use

7.1 Contacting and contributing to SPASE

The NASA-based SPASE efforts now have an infrastructure that maintains and updates the Data Model along with needed documents and tools with input from all of the SPASE Consortium. The inclusion of new projects and products into the SPASE Registry is not yet automated and will always involve some checking by a person to make sure everything is in order. Those needing help with creating or editing SPASE documents should contact us (see https://hpde.gsfc.nasa.gov/spase_metadata.html). To see examples of SPASE descriptions, use the Heliophysics Data Portal (see Sec. 7.3 below).

7.2 SPASE editor

The generation of SPASE descriptions is, in principle, relatively simple at the level of the required entries (ID, Header, Measurement Type, Temporal Description, and a few others). Even that level can be enough work to be an impediment to implementation. Various experts have provided most of the descriptions to date, using prior cases as starting points or tools such as “ADAPT” that cull metadata from Common Data Format (CDF) data files for the product and other sources. As an example, here is what the beginning of an XML file describing a Numerical Data Resource looks like:


```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Spase xmlns="http://www.spase-group.org/data/schema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.spase-group.org/data/schema  http://www.spase-
group.org/data/schema/spase-2_3_0.xsd">
  <Version>2.3.0</Version>
  <NumericalData>
    <ResourceID>spase://JAXA/NumericalData/Geotail/MGF/PT15S</ResourceID>
    <ResourceHeader>
      <ResourceName>Geotail 15-sec magnetic field data, solar wind
             only</ResourceName>
      <ReleaseDate>2018-06-28T22:02:01Z</ReleaseDate>
      <Description>Data consist of 15-sec averages of magnetic field
magnitude and GSE Cartesian components, from the MGF magnetometer on
Geotail.  Data are for only the solar wind phases of the Geotail orbit.
The Geotail position vector in GSE coordinates is included.</Description>
      <Acknowledgement>Dr. T. Nagai</Acknowledgement>
      <Contact>
        <PersonID>spase://SMWG/Person/Tsugunobu.Nagai</PersonID>
        <Role>PrincipalInvestigator</Role>
```

```
      </Contact>
      <Contact>
        <PersonID>spase://SMWG/Person/Natalia.E.Papitashvili</PersonID>
        <Role>DataProducer</Role>
      </Contact>
      <InformationURL>
        <Name>Readme file at SPDF</Name>
        <URL>ftp://spdf.gsfc.nasa.gov/pub/data/
                         geotail/mgf/mag_sw_15s_ascii/00readme</URL>
        <Description>Details on creation of this data set</Description>
      </InformationURL>
      <PriorID>spase://VMO/NumericalData/Geotail/MGF/PT15S</PriorID>
      <PriorID>spase://VSPO/NumericalData/Geotail/MGF/PT15S</PriorID>
    </ResourceHeader>
…
```

In the interest of making the process simple, requiring no knowledge of XML and the "schema" used to insure they are correctly formatted (see https://www.w3schools.com/xml/schema_example.asp for the use of schema), we are constructing an online editor. Figure 4 shows a snapshot of the editor that will take care of the structure of SPASE and just ask the user for the terms and descriptions that make the product useful. When a term is part of an explicit list, the user is shown a list from which to choose. This editor will be announced to the HP community when a beta-test version is available. As mentioned above, there is already a tool available for checking the validity of independently generated SPASE descriptions (http://spase-group.org/tools/validate/).

7.3 Finding, Plotting, and Accessing HP Products

The uniformity of the SPASE metadata makes it easy to create tools to find particular data products. The Heliophysics Data Portal (https://heliophysicsdata.gsfc.nasa.gov) is such a search tool (Figure 5). It is a public face for the inventory, serving as a "card catalog" that also can deliver the "books" (data). The HDP reads the SPASE inventory into a database that allows searching using many SPASE terms or time- and text-based searches to find HP data products. In Fig. 4, the left column has boxes for text and time searches ("restrictions on the product list"), and a list of categories that can be used for further restrictions. At any point, the full list of resources consistent with the list of restrictions may be viewed: a "View Current List" link will appear at the top of the right column if more than 20 products in the list. The list can be sorted by Observatory, Cadence, Measurement Type, etc. Once the user finds a suitable product, the right columns provide descriptions based on SPASE and Access Links to the data and documentation. "Get Data" buttons provide links to plots and data from CDAWeb and other sources.

7.4 Application Programmer Interfaces for remote machine access to data

As part of SPASE descriptions, “Product Key” and “Parameter Keys” (indicated in purple in Fig. 1) provide a unique product identifier within the serving repository and the variable identifiers required to request subsets of data using direct calls to data servers rather than through a Web portal. This allows a user to, for example, directly get data from an IDL or Python session using simple commands that construct calls to Web Services. A uniform specification of the behavior of data servers, with corresponding client (user) software to communicate with the servers, makes it possible for any data supplier to provide data with uniform calling methods. Since it is the response of the server to particular requests that is specified, any appropriate language can be used (Java, Python, etc.) to implement the “Application Programmer Interface” (API). Thus, the syntax for a simple Heliophysics API (HAPI; see https://github.com/hapi-server) call for data is:

```
http://hapi-server.org/hapi/data?id=path/to/ACE_MAG
        &time.min=2016-01-01T00:00:00.000Z
        &time.max=2016-02-01T13:10:30.000Z&include=header
```

where “ACE_MAG” is the Product Key. In this case, all variables would be delivered. Using the Parameter Keys “Bx” and “By” would yield a subset of the data containing only Bx and By:

```
http://hapi-server.org/hapi/data?id=MY_MAG_DATA
        &parameters=Bx,By&time.min=1999-06-03Z&time.max=2000-07-22Z
```

The following URL (without the line breaks) will deliver data when entered into a browser or when generated by an application:

```
https://cdaweb.gsfc.nasa.gov/hapi/data?id=AC_H1_MFI
    &time.min=2006-01-01
         &time.max=2006-01-03&parameters=Magnitude,BGSEc
```

Client software just needs to generate and submit URLs of the above syntax to any HAPI server to get the requested information in response. Usually the data would be directly loaded into the user application using software that “understands” the (uniform) HAPI data format.

There are a number of APIs for different systems including, importantly, that for the Virtual Solar Observatory (https://vso.nascom.nasa.gov/API/, https://sdac.virtualsolar.org/), which is now used extensively for solar data retrieval, and “CDAS”, which is used for the retrieval of data from NASA’s Space Physics Data Facility (SPDF) “CDAWeb” archive (https://cdaweb.gsfc.nasa.gov/WebServices/, https://cdaweb.gsfc.nasa.gov). While there are efforts on-going to obtain wide acceptance of HAPI as a simple, universal web service, it is not expected that the older APIs will lose their utility.

7.5 Present and future space weather applications

All the applications of SPASE discussed above are directly applicable to space weather research. One key to such research is the ability to use a multiplicity of datasets for a given project, and this is what SPASE is designed for. Combined with other standards, SPASE facilitates an integrated approach in which any required data can be quickly found and accessed. Current efforts will expand these capabilities by, for example, making it easier to run and validate models by keeping track of all parts of a data/simulation chain (see above on CCMC). Near future plans include incorporating real-time data into the SPASE framework using connections to iSWA and to the Space Weather Prediction Center (https://www.swpc.noaa.gov). We plan to use the HAPI protocol to standardize these connections, and this will facilitate bringing real-time data into common graphics and analysis packages.

7.6 Linking with other developments

The Solar System science community (thus including the Solar, Earth and Planetary Magnetosphere science topics) is building interoperable frameworks for finding, accessing and reusing data. A major initiative is the Europlanet-2020/VESPA project (Virtual European Solar and Planetary Access) (Erard et al., 2017). VESPA is providing an infrastructure to search for data products based on science content and coverage query parameters. This is enabled through a data model called EPNcore (Europlanet core metadata) (Erard et al., 2017), coupled with the Table Access Protocol (Dowler et al., 2018) developed and maintained by the IVOA. Many HP related data collections are available through the VESPA query interfaces (either the main VESPA query portal, http://vespa.obspm.fr, or from interfaces embedded in tools, such as in AMDA (Automated Multi-Dataset Analysis; http://amda.irap.omp.eu from the CDPP)). The team developing the generic plotting and analysis "Autoplot" application (http://autoplot.org) is working with the VESPA team to be able to connect with the VESPA infrastructure (e.g., with the Simple Application Messaging Protocol (SAMP) (Taylor et al., 2012)). The MASER (Measuring, Analyzing and Simulating Radio Emissions) team is aiming at sharing low frequency radio data using existing infrastructures (Cecconi et al., 2018). The mapping between the relevant EPNcore and SPASE dictionary elements has already been done by the AMDA team for their internal database. Current developments in the VESPA team include working on using VESPA as a searchable registry for HP webservices, including IMPEx, HAPI and "Das2" (Piker et al., 2017) interfaces. The webservices would then be discoverable with data collections from HP and neighboring fields such as planetary aurorae, e.g., with the APIS (Auroral Planetary Imaging and Spectroscopy) (Lamy et al., 2015), which implements a VESPA search interface. The publication of the SPASE resource tree in VESPA would enhance the visibility of the data products and collections. The bridging with the astronomy community standards and tools is also being done by the Solar System Interest Group of IVOA, whose goal is to adapt IVOA standards to the needs of Solar System science. In one other development involving SPASE, the ESA Space Situational Awareness (SSA) pre-operational space weather services development is currently assessing improved standardization of metadata and terminologies. The working group involved with this is strongly considering the use of SPASE due to its robust heritage, widespread use and ongoing active support.

A more detailed view of the operation of AMDA and its use of SPASE is shown in Figure 6. In AMDA, data are presented in a window called a "Workspace Explorer" as a hierarchy of missions, instruments, datasets and parameters. Missions may be space missions as well as ground based observatories or simulations. These concepts are mapped to SPASE resources: Observatory, Instrument, and NumericalData. The SPASE XML files are stored in a Registry and used to display information at each level in the Workspace Explorer, after transformation using a style sheet. The main component in AMDA is a SPASE parameter. SPASE Parameters are used to access the actual data in the Plot Manager. SPASE metadata are also displayed as titles for axes on the plot. A user selects a physical quantity in the Workspace Explorer that provides SPASE-based information on the corresponding mission, instrument, and dataset. The selected parameter is dragged and dropped to the Plot Manager in which it can be plotted over a selected time interval.

The same Registry that is used in AMDA is also used for VESPA. The EPN-TAP protocol allows access to granules, which are typically files. Each file is described in a row of the relational database, and the SPASE XML Registry is used to populate the rows, after a translation from SPASE to EPNCore. Some metadata not defined in EPNCore are added as optional for EPN-TAP. For example, 'observed_region' (SPASE) is translated to 'spase_region' (EPNCore). Data from AMDA may be searched from VESPA using the common Registry. Users searching for data of interest through e.g. the VESPA portal (see above) may choose several search criteria, most of them compatible with SPASE (instrument, timespan...) and get a list of files (one file per row) with a link to access the actual data. The same operations are possible with other VESPA clients like 3DView. The connection of the VESPA service to AMDA is depicted in Figure 7.

7.7 Digital Object Identifiers for Data

SPASE Product IDs provide unique identifiers of HP data products. This is precisely what is required for the citation of data products in what is becoming the standard approach to the recognition of data in most journals. Thus, the SPASE group has begun to issue Digital Object Identifiers for datasets, starting with NASA mission products. The mission teams determine the appropriate "Author," "Publisher," and "Publication Date," and the SPASE group uses its UCLA connection to DataCite for DOI minting. This information is then added to the SPASE record for a given product and becomes a link to a doi.org landing page, which is, in turn, based on the SPASE description. The DOI becomes the permanent link to the dataset, allowing references to datasets in the same way as to published papers. Non-NASA data product DOIs will be arranged with the organizations and people associated with each dataset, but the SPASE group, among many others (see https://doi.org) can be used to obtain DOIs. The SPASE registry provides data products at the right level for issuing DOIs, and the requirement of having DOIs for all data will ensure precise product distinctions in cases where the current descriptions are not adequate. Note that there is flexibility in the definition of a Data Product for the assignment of DOIs, and in particular a single DOI can be used for a dataset that is current, with records being added as new measurements are made. The data providers are responsible for determining how much of a “version” change merits a new DOI, although changes such as the adding of parameters would lead to a new Product, and thus a new DOI.

8 Conclusions

The SPASE Data Model has grown over the last decade from early ideas stating that a unified approach to data discovery, access, and use is how we will make progress on current HP problems to a rich metadata system that facilitates many tasks, from data citation with DOIs to data access with APIs. The model is stable but growing with the changes in the demands put on it. NASA has affirmed that the effort should be supported as an infrastructure project, and this assures longevity. There are still challenges to more fully integrating SPASE descriptions into, for example, a HAPI framework in which they could provide an automated source of required metadata. While the full implementation of SPASE metadata into end-to-end simulations of, e.g., space weather events will be a challenge for some time, the many applications outlined above will make it progressively more useful. Community-wide adoption of the SPASE Data Model as a metadata standard will enable the development of interoperable data tools and a more effective data environment.

Acknowledgments

Listed US Authors (except CW) acknowledge the support of NASA’s Heliophysics Data and Model Consortium. CW is supported by the CCMC. The Centre de Données de la Physique des Plasmas (CDPP) is supported by CNRS, CNES, the Observatoire de Paris and Université Paul Sabatier, Toulouse; the CDPP provides support for VG, MG, and BC. The IMPEx functionalities and SH, VG, and MG have been supported by the EU FP7 project “IMPEx” under grant agreement No 262863. The "Europlanet H2020 Research Infrastructure" project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 654208; VG, MG, and BC received support from this source. MG receives support from the ASOV (Action Spécifique Observatoire Virtuel) whose activities are funded by CNRS/INSU. BC received support from the IPDA (International Planetary Data Alliance) whose activities are funded by CNES. We also thank the Editors and referees for many helpful comments.

References

Bogart, R., Hanisch, R, J. King, J., Pyle, R., Sibeck, D., Walker, R. & Winningham, D. (1998). *Report of the Sun-Earth Connection Study Team of the Space Science Data Systems Technical Working Group*. https://cdaweb.gsfc.nasa.gov/pub/documents/HPDE/Riverbend_Report.doc

Cecconi B, Le Sidaner P, Savalle R, Bonnin X, Zarka P, Louis C, Coffre A, Aicardi S, Lamy L, Denis L, Grießmeier J-M, Faden J, Piker C, André N, Génot V, Erard S, Mafi J N, King T A, Sharlow M, Sky J, Demleitner M, (2018) MASER: A Toolbox for Low Frequency Radio Astronomy, *PV-2018 proceedings*, 20–24, http://purl.org/net/epubs/manifestation/37981071

Dowler, P, Rixon, G, Tody, D, Demleitner, M, (2018). Table Access Protocol, Version 1.1, *IVOA Documentation*, http://www.ivoa.net/documents/TAP/

Erard, S., Cecconi, B., Le Sidaner, P., Rossi, A. P., Capria, M. T., Schmitt, B, Génot, V., André, et al. (2017) VESPA: A community-driven Virtual Observatory in Planetary Science. *Planetary and Space Science*, 150, 65–85., https://doi.org/10.1016/j.pss.2017.05.013

Génot, V., Beigbeder, L., Popescu, D., Dufourg, N., Gangloff, M., Bouchemit, M., Bouchemit, M., Caussarieu, S., Toniutti, J.-P., Durand, J., Modolo, R., André, N., Cecconi, B., Jacquey, C., Pitout, F., Rouillard, A., Pinto, R., Erard, S., Jourdane, N., Leclercq, L., Hess, S., Khodachenko, M., Al-Ubaidi, T., Scherf, M., Budnik, E. (2018). Science data visualization in planetary and heliospheric contexts with 3DView. *Planetary and Space Science*, *150*, 111–130. https://doi.org/10.1016/j.pss.2017.07.007

Harvey, C. C., M. Gangloff, King, T., Perry, C.H., Roberts, D. A., & Thieman, J. (2008). Virtual Observatories for space and solar physics research, *Earth Science Informatics*, 2008, https://doi.org/10.1007/s12145-008-0008-1.

King, T., Thieman, J., and Roberts, D. A. (2010). SPASE 2.0: A Standard Data Model for Space Physics, *Earth Science Informatics*, 3:67–73. https://doi.org/10.1007/s12145-010-0053-4

Lamy, L., Prangé, R., Henry, F., & Le Sidaner, P. (2015). The Auroral Planetary Imaging and Spectroscopy (APIS) service. *Astronomy and Computing*, *11*, 138–145. https://doi.org/10.1016/j.ascom.2015.01.005

NASA Heliophysics Division Director, HPDD (2016). *NASA Heliophysics Scientific Data Management Policy*, Version 1.2, NASA, https://doi.org/10.21978/P83P78

National Research Council, Space Studies Board Committee on Data Management and Computation ("CODMAC") (1982). Data Management and Computation–Volume I: Issues and Recommendations. Washington, DC: *The National Academies Press*. https://doi.org/10.17226/12366

Piker, C, et al. Das2 server ICD, (2017). *University of Iowa*. http://das2.org/Das2.2.2-ICD_2017-05-09.pdf

Taylor, M., Boch T., Fitzpatrick, M., Allan, A., Fay, J., Paioro L., Taylor, J., Tody, D., (2012) Simple Application Messaging Protocol, Version 1.3, *IVOA Documentation*. http://ivoa.net/documents/SAMP/

The SPASE Consortium and IMPEx (2014) *Simulation Extensions for the Space Physics Archive Search and Extract (SPASE) Data Model, Version 1.0.0*, http://spase-group.org/docs/dictionary/spase-sim-1_0_0.pdf

The SPASE Consortium (2018). *A Space and Solar Physics Data Model from the SPASE Consortium, Version 2.3.0*, http://spase-group.org/docs/dictionary/spase-2_3_0.pdf

Thieman, J. R., Roberts, D. A., King, T. A., Harvey, C. C, Perry, C, & Richards, P. J. (2010). SPASE and the Heliophysics Virtual Observatories. *Data Science Journal*, 9, https://doi.org/10.2481/dsj.SS_IGY-019

Figure Captions

Figure 1. Overview of SPASE, showing the Resource Types in the left column and the (partial) hierarchy of the description of Numerical Data and Instrument Resources to the right of that. The full structure of SPASE can be found in The SPASE Consortium (2014, 2018).

Figure 2. A diagrammatic way of showing the full set of entries in a SPASE Resource Header. The SPASE site contains a complete set of such descriptors.

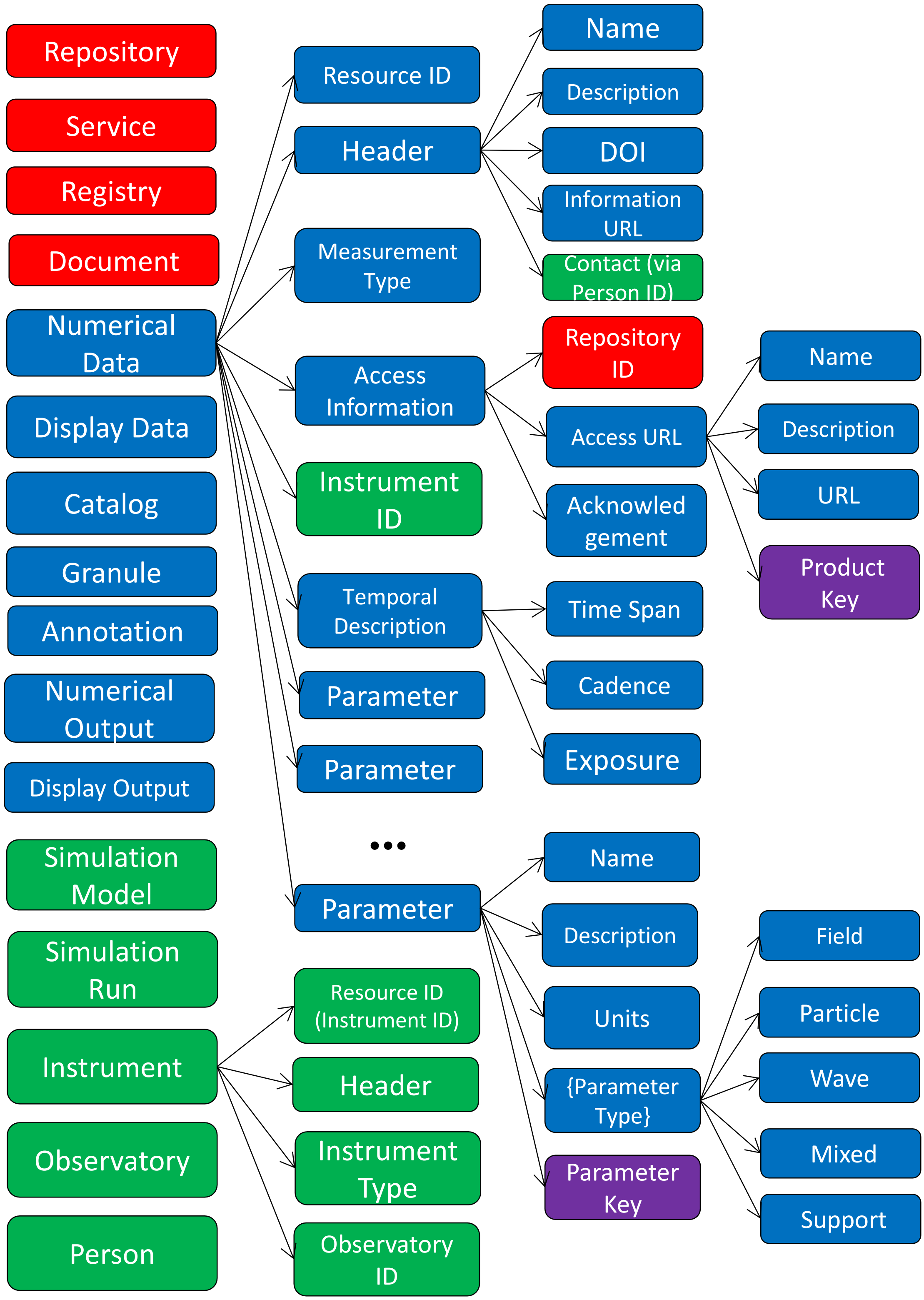
Figure 3. An overview of the role of SPASE in the CCMC. The CMA is a database view of the SPASE metadata associated with CCMC models and runs. It can export SPASE XML to allow other services to interact with the simulations, thus expediting data-model comparisons.

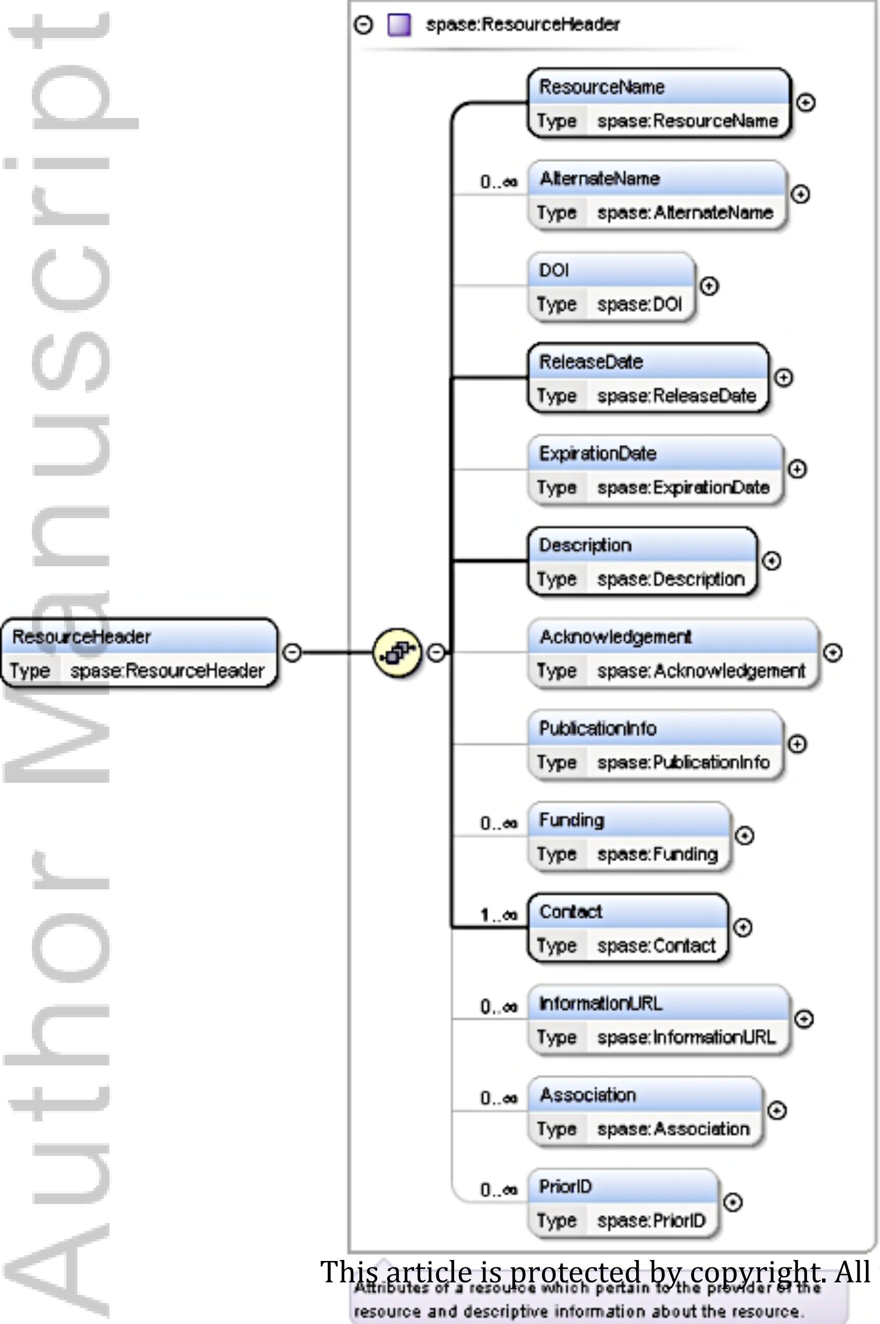
Figure 4. A sample page from an online SPASE resource editor. The user only needs to know what the product is, not the details of XML. (Use the online version of the figure to magnify the view and thus the text. This editor is to be online soon, at http://spase-group.org.)

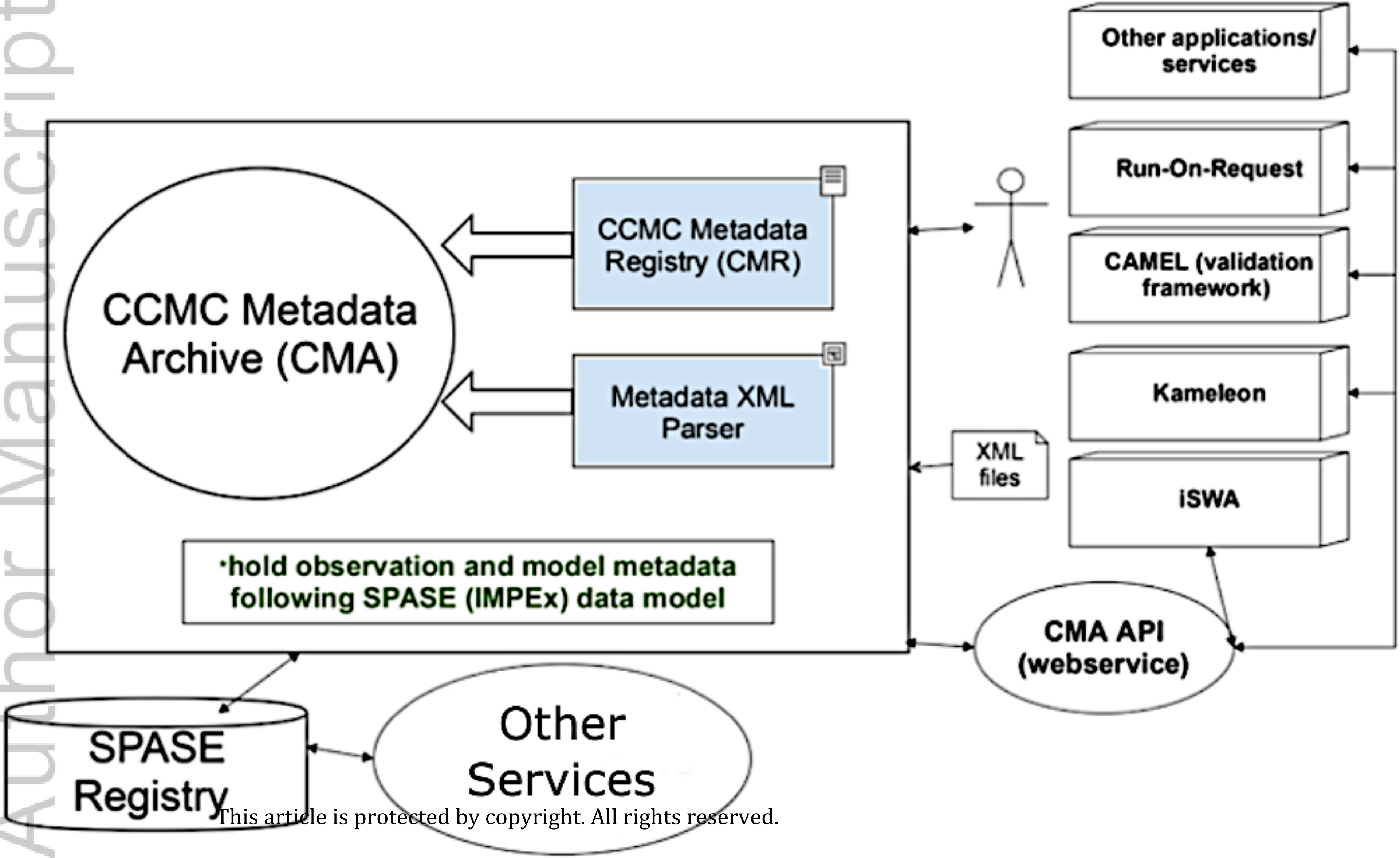
Figure 5. An instance of a search using the Heliophysics Data Portal that serves as a “public face” of the SPASE inventory. Shown here is a partial listing for a search for interplanetary plasma data in 1976.

Figure 6. The use of SPASE in AMDA. The uniform metadata aids product discovery and understanding, parameter (variable) references and retrieval, and variable naming on plots. (Use the online version of the figure to magnify the view, and thus the text.)

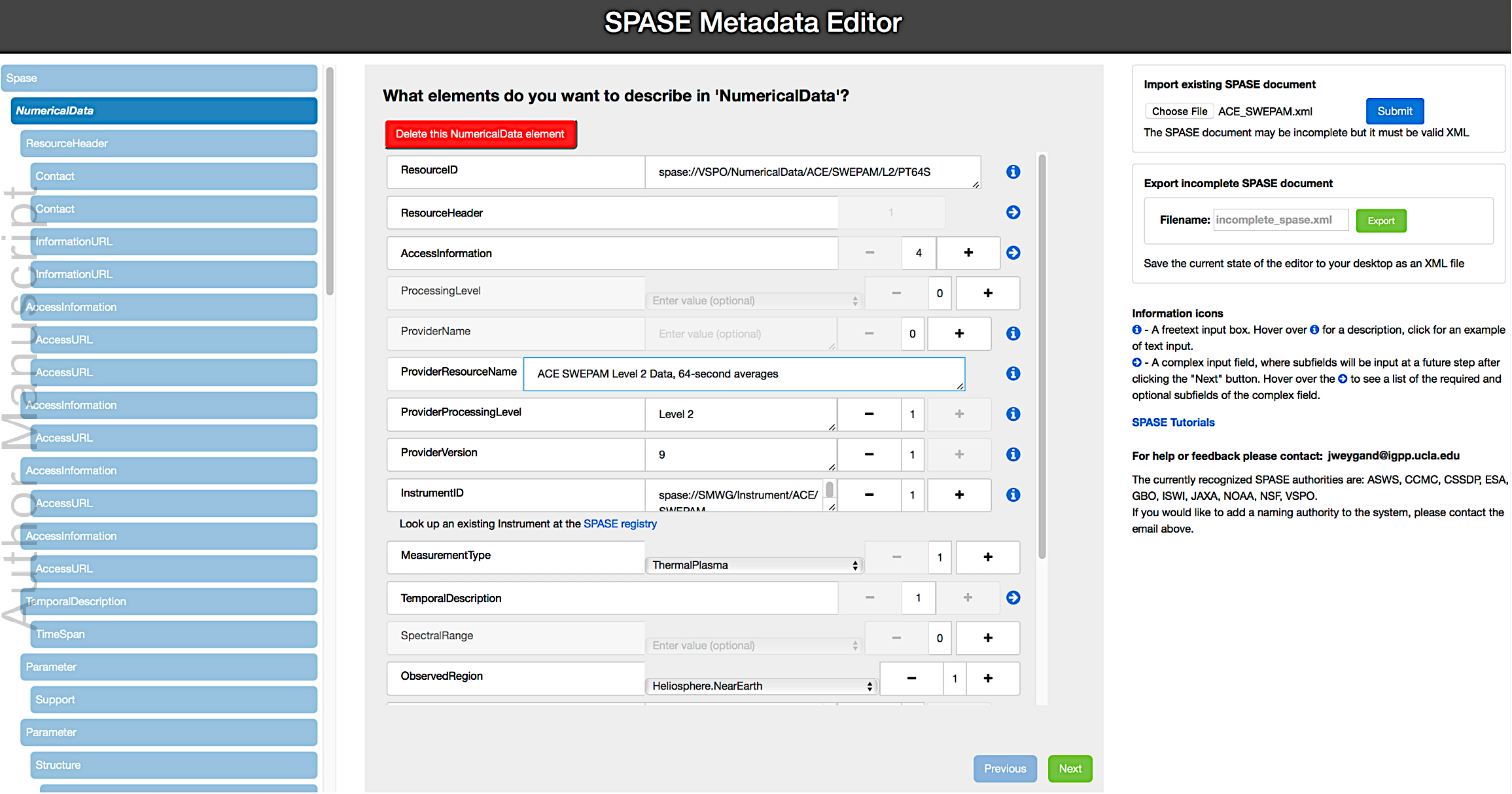
Figure 7. Metadata connection between AMDA and VESPA. Note that a SPASE-compliant Registry is key to the linkage.

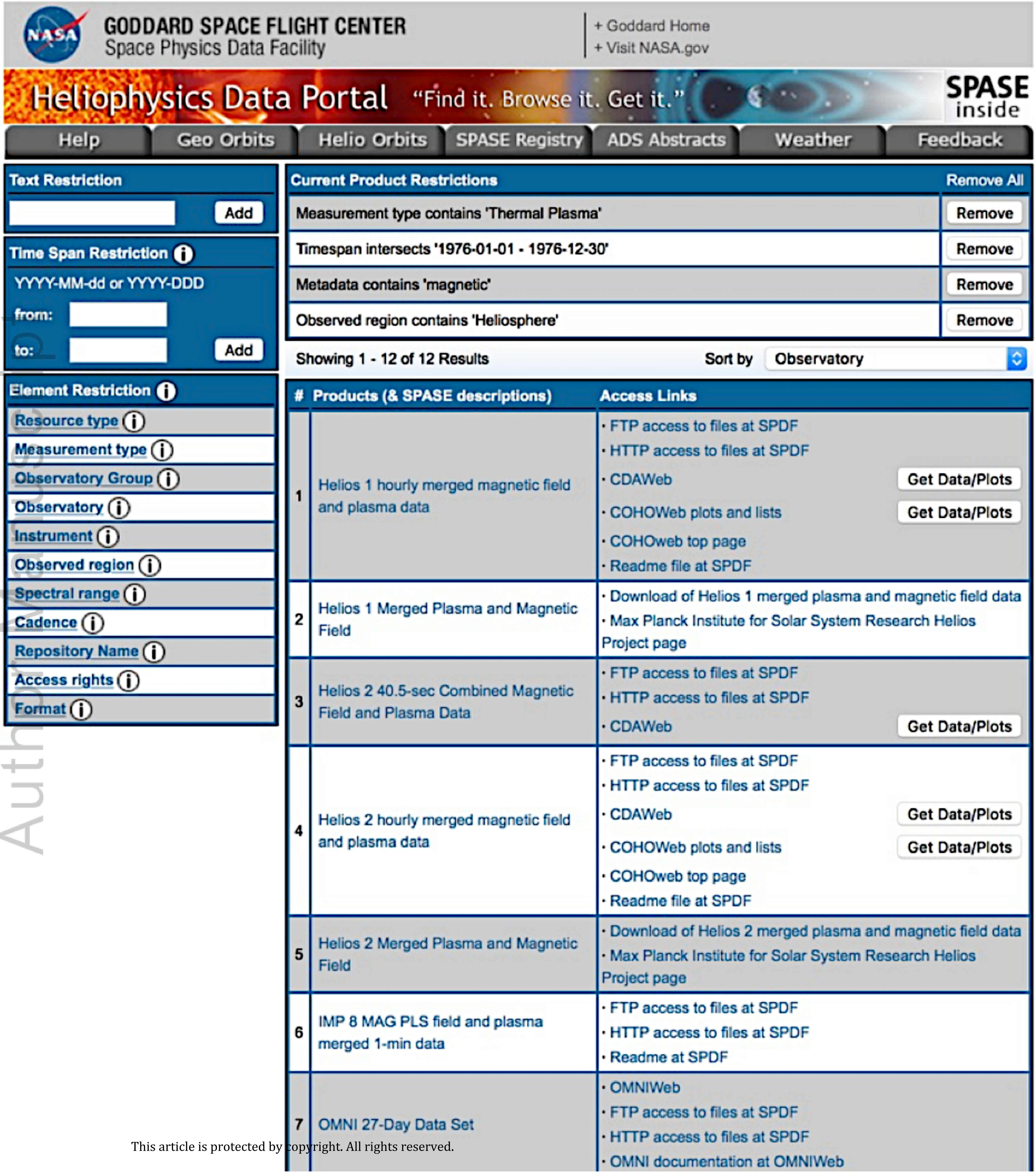

Repository
Service
Registry
Document
Numerical Data
Display Data
Catalog
Granule
Annotation
Numerical Output
Display Output
Simulation Model
Simulation Run
Instrument
Observatory
Person
Resource ID
Header
Measurement Type
Access Information
Instrument ID
Temporal Description
Parameter
Parameter
•••
Parameter
Resource ID (Instrument ID)
Header
Instrument Type
Observatory ID
Name
Description
DOI
Information URL
Contact (via Person ID)
Repository ID
Access URL
Acknowledgement
Time Span
Cadence
Exposure
Name
Description
Units
{Parameter Type}
Parameter Key
Name
Description
URL
Product Key
Field
Particle
Wave
Mixed
Support

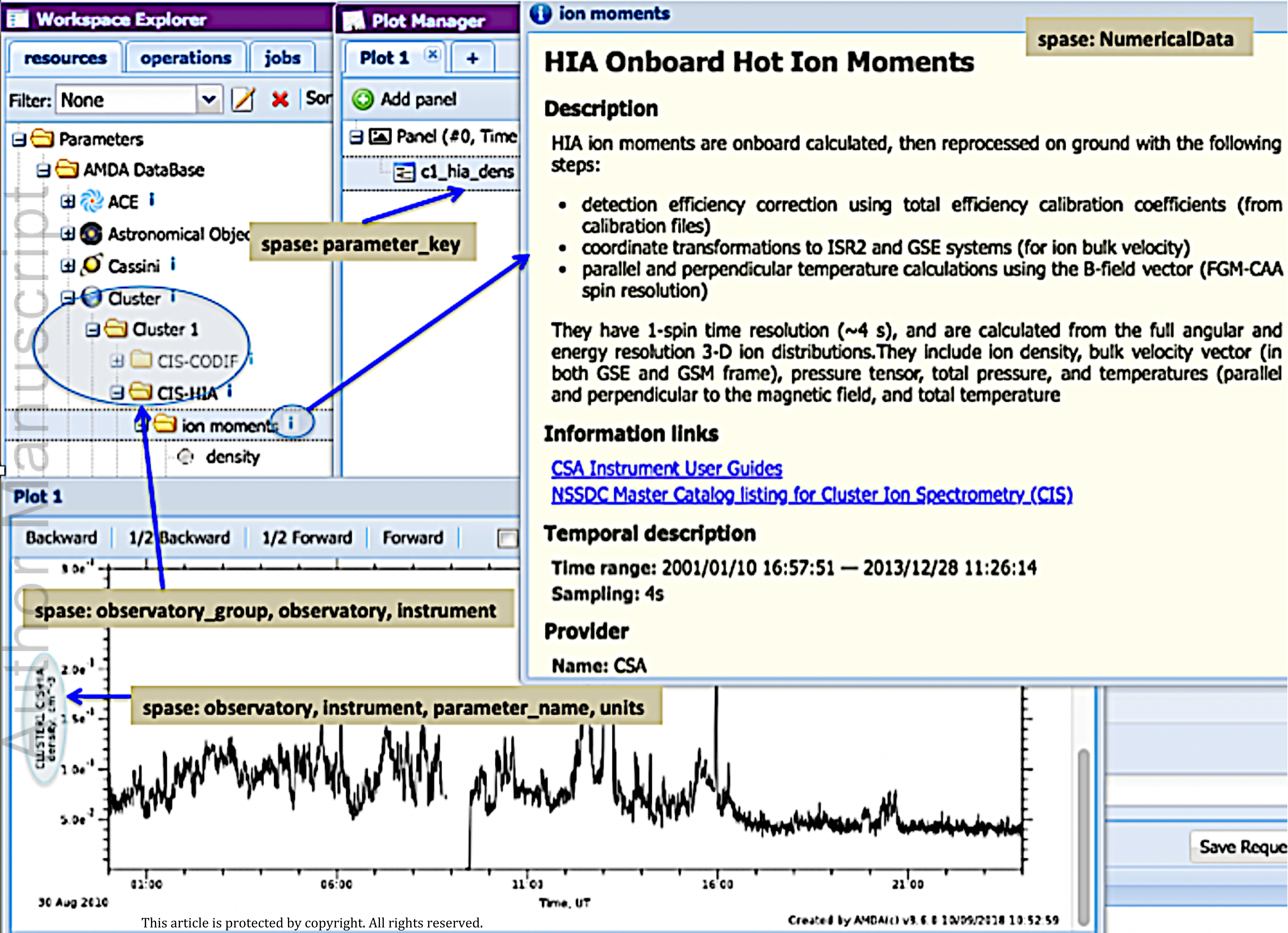

spase:ResourceHeader
ResourceHeader
Type spase:ResourceHeader
ResourceName
Type spase:ResourceName
0..∞
AlternateName
Type spase:AlternateName
DOI
Type spase:DOI
ReleaseDate
Type spase:ReleaseDate
ExpirationDate
Type spase:ExpirationDate
Description
Type spase:Description
Acknowledgement
Type spase:Acknowledgement
PublicationInfo
Type spase:PublicationInfo
0..∞
Funding
Type spase:Funding
1..∞
Contact
Type spase:Contact
0..∞
InformationURL
Type spase:InformationURL
0..∞
Association
Type spase:Association
0..∞
PriorID
Type spase:PriorID
Attributes of a resource which pertain to the provider of the resource and descriptive information about the resource.

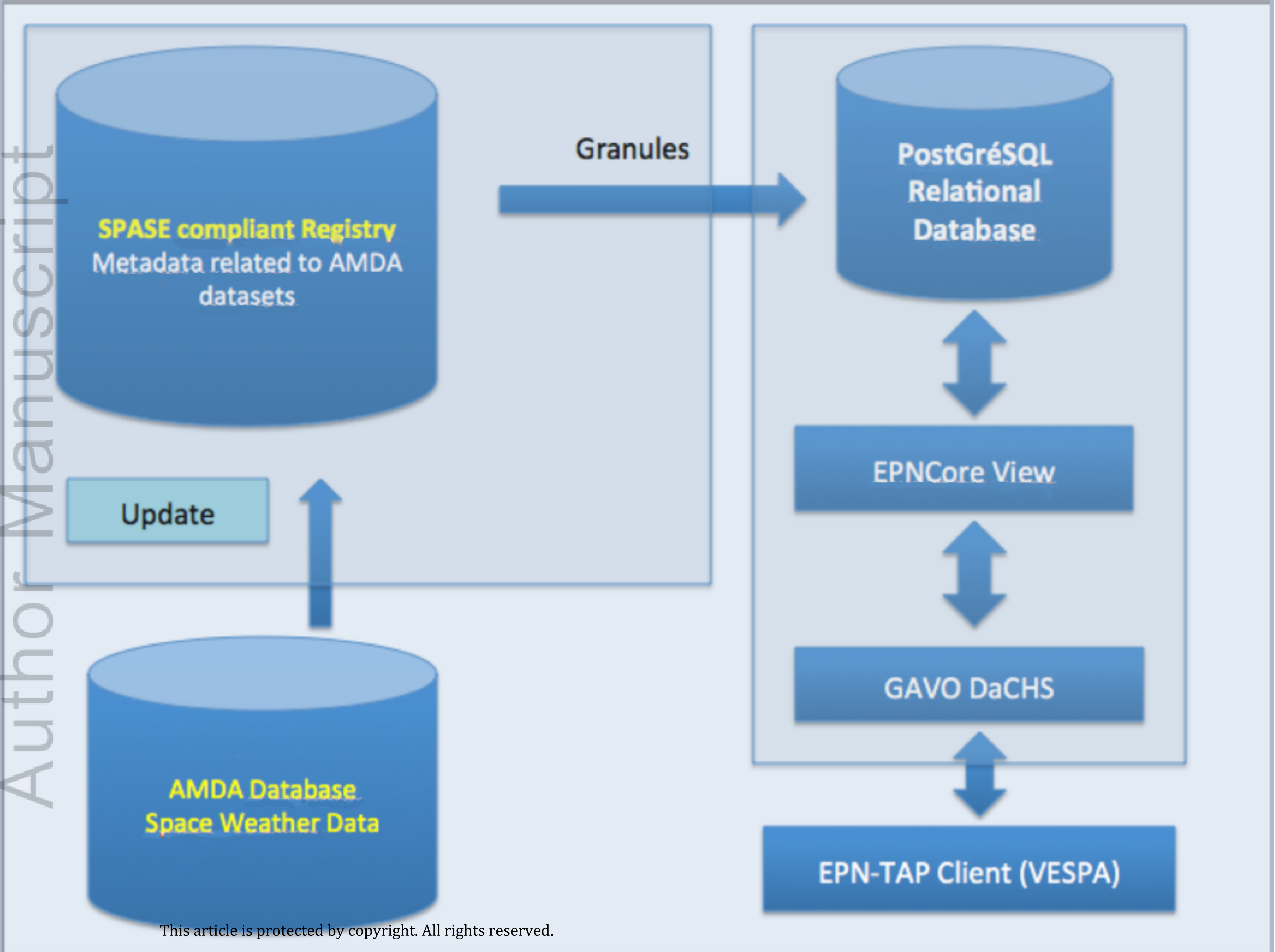


Other applications/ services
Run-On-Request
CAMEL (validation framework)
Kameleon
iSWA
CCMC Metadata Archive (CMA)
CCMC Metadata Registry (CMR)
Metadata XML Parser
XML files
·hold observation and model metadata following SPASE (IMPEx) data model
CMA API (webservice)
SPASE Registry
Other Services


SPASE Metadata Editor
Spase
NumericalData
ResourceHeader
Contact
Contact
InformationURL
InformationURL
AccessInformation
AccessURL
AccessURL
AccessInformation
AccessURL
AccessInformation
AccessURL
AccessInformation
AccessURL
TemporalDescription
TimeSpan
Parameter
Support
Parameter
Structure
What elements do you want to describe in 'NumericalData'?
Delete this NumericalData element
ResourceID
spase://VSPO/NumericalData/ACE/SWEPAM/L2/PT64S
ResourceHeader
AccessInformation
ProcessingLevel
Enter value (optional)
ProviderName
Enter value (optional)
ProviderResourceName
ACE SWEPAM Level 2 Data, 64-second averages
ProviderProcessingLevel
Level 2
ProviderVersion
9
InstrumentID
spase://SMWG/Instrument/ACE/
Look up an existing Instrument at the SPASE registry
MeasurementType
ThermalPlasma
TemporalDescription
SpectralRange
Enter value (optional)
ObservedRegion
Heliosphere.NearEarth
Previous
Next
Import existing SPASE document
Choose File
ACE_SWEPAM.xml
Submit
The SPASE document may be incomplete but it must be valid XML
Export incomplete SPASE document
Filename:
incomplete_spase.xml
Export
Save the current state of the editor to your desktop as an XML file
Information icons
- A freetext input box. Hover over for a description, click for an example of text input.
- A complex input field, where subfields will be input at a future step after clicking the "Next" button. Hover over the to see a list of the required and optional subfields of the complex field.
SPASE Tutorials
For help or feedback please contact: jweygand@igpp.ucla.edu
The currently recognized SPASE authorities are: ASWS, CCMC, CSSDP, ESA, GBO, ISWI, JAXA, NOAA, NSF, VSPO.
If you would like to add a naming authority to the system, please contact the email above.


GODDARD SPACE FLIGHT CENTER
Space Physics Data Facility
+ Goddard Home
+ Visit NASA.gov
Heliophysics Data Portal
"Find it. Browse it. Get it."
SPASE inside
Help
Geo Orbits
Helio Orbits
SPASE Registry
ADS Abstracts
Weather
Feedback
Text Restriction
Add
Time Span Restriction
YYYY-MM-dd or YYYY-DDD
from:
to:
Add
Element Restriction
Resource type
Measurement type
Observatory Group
Observatory
Instrument
Observed region
Spectral range
Cadence
Repository Name
Access rights
Format
Current Product Restrictions
Remove All
Measurement type contains 'Thermal Plasma'
Remove
Timespan intersects '1976-01-01 - 1976-12-30'
Remove
Metadata contains 'magnetic'
Remove
Observed region contains 'Heliosphere'
Remove
Showing 1 - 12 of 12 Results
Sort by
Observatory
Products (& SPASE descriptions)
Access Links
1
Helios 1 hourly merged magnetic field and plasma data
· FTP access to files at SPDF
· HTTP access to files at SPDF
· CDAWeb
Get Data/Plots
· COHOWeb plots and lists
Get Data/Plots
· COHOweb top page
· Readme file at SPDF
2
Helios 1 Merged Plasma and Magnetic Field
· Download of Helios 1 merged plasma and magnetic field data
· Max Planck Institute for Solar System Research Helios Project page
3
Helios 2 40.5-sec Combined Magnetic Field and Plasma Data
· FTP access to files at SPDF
· HTTP access to files at SPDF
· CDAWeb
Get Data/Plots
4
Helios 2 hourly merged magnetic field and plasma data
· FTP access to files at SPDF
· HTTP access to files at SPDF
· CDAWeb
Get Data/Plots
· COHOWeb plots and lists
Get Data/Plots
· COHOweb top page
· Readme file at SPDF
5
Helios 2 Merged Plasma and Magnetic Field
· Download of Helios 2 merged plasma and magnetic field data
· Max Planck Institute for Solar System Research Helios Project page
6
IMP 8 MAG PLS field and plasma merged 1-min data
· FTP access to files at SPDF
· HTTP access to files at SPDF
· Readme at SPDF
7
OMNI 27-Day Data Set
· OMNIWeb
· FTP access to files at SPDF
· HTTP access to files at SPDF
· OMNI documentation at OMNIWeb


Workspace Explorer
resources
operations
jobs
Filter: None
Parameters
AMDA DataBase
ACE
Cassini
Cluster
Cluster 1
CIS-CODIF
CIS-HIA
ion moments
density
Plot Manager
Plot 1
Add panel
Panel (#0, Time
c1_hia_dens
ion moments
spase: NumericalData
HIA Onboard Hot Ion Moments
Description
HIA ion moments are onboard calculated, then reprocessed on ground with the following steps:
• detection efficiency correction using total efficiency calibration coefficients (from calibration files)
• coordinate transformations to ISR2 and GSE systems (for ion bulk velocity)
• parallel and perpendicular temperature calculations using the B-field vector (FGM-CAA spin resolution)
They have 1-spin time resolution (~4 s), and are calculated from the full angular and energy resolution 3-D ion distributions.They include ion density, bulk velocity vector (in both GSE and GSM frame), pressure tensor, total pressure, and temperatures (parallel and perpendicular to the magnetic field, and total temperature
Information links
CSA Instrument User Guides
NSSDC Master Catalog listing for Cluster Ion Spectrometry (CIS)
Temporal description
Time range: 2001/01/10 16:57:51 — 2013/12/28 11:26:14
Sampling: 4s
Provider
Name: CSA
spase: parameter_key
spase: observatory_group, observatory, instrument
spase: observatory, instrument, parameter_name, units
Plot 1
Backward
1/2 Backward
1/2 Forward
Forward
Save Reque
Time, UT
30 Aug 2010
Author Manuscript


Granules
SPASE compliant Registry
Metadata related to AMDA datasets
Update
AMDA Database
Space Weather Data
PostGréSQL Relational Database
EPNCore View
GAVO DaCHS
EPN-TAP Client (VESPA)