Recent developments towards a Solar System Virtual Observatory

C.C. Harvey, M. Gangloff, C. Jacquey (CNRS/CDPP)
J.R. Thieman (NASA/NSSDC), D.A. Roberts (NASA/GSFC)
T. King (UCLA/PDS)

Outline

- The Space Physics Data Environment

- What is SPASE ?

- The SPASE Data Model

- Future users

- European Considerations

- Virtual Observatories and Space Weather

- Future Plans

The Space Physics Data Environment

- Thousands of relatively small datasets as well as several large datasets

- Tens to hundreds of data centers or data providers (repositories), scattered worldwide

- Very loose (if any) coupling of data centers

- Very diverse metadata terminology and data formats

- Data which is not archived will effectively be lost.

- Preservation of the data often becomes the task of a national data centre (NSSDC, CDPP)

Recent Space and Solar Physics Missions

Historical Background

- Data centres were set up to aggregate, preserve and promote the use of data, and they set there own standards about how best to do this.
- The requirement to search worldwide across multiple data centres has long been recognised, and in 1998 ISTP asked four data centres to suggest a solution.
- Informal discussion continued until …
- In October 2003 an international consortium, SPASE, was formally created by these and other data centres.
- The worldwide scene continues to evolve until today the players now are …

The Players

- **<u>Virtual observatories</u>** provide uniform access to data and services for some particular group of users, whose contours are defined by considerations of scientific interest, resources, and geo-politics.
- **<u>Data repositories</u>** preserve data and make it accessible over the Internet. They vary from instrument specific to broadly based, and
  - are generally supported by national or international agencies, and
  - are accessed either directly or via one or more of the virtual observatories.
- **<u>Service centres</u>** supply or make available services (*e.g.*, SSC, CoSEC,).
- **<u>Data (or service) providers</u>** supply the data (or service software) made available via some data repository (or service centre).
- **<u>End users</u>** are the scientists in the field who submit their user requirements to the virtual observatories
- **<u>SPASE</u>** promotes interoperability within & between VOs serving Solar System physics research, by elaboration of standards & implement-ation guidelines. SPASE provides no infrastructure or other resource.

What is SPASE ?

Space Physics Archive Search and Extract is :

- 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

- Formalised in 2003 as an international consortium with an open invitation for anyone in the community to participate actively

- U.S. participants funded by NASA since July 2005. SPASE will receive "permanent support" from NSSDC after 2007.

- SPASE is the only international instance supporting global data management for Solar and Space Physics

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.

This will enable ...

- One-stop searching for data in multiple data repositories

- Intercomparison of similar quantities from different data sets in different data centres through common terminology mapping (*e.g.* visible radiation *vs.* optical radiation)

- Identification of disparate data granules overlapping in time.

- Recovery of data granules together with all the information necessary for their immediate analysis.
(Realisation of this long-term objective depends upon the ability of VOs to provide suitably formatted data and metadata)

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 discussion of a Space Physics Data Model and Data Dictionary.
- Fully discuss (e-mail + fortnightly teleconferences) 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 the 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 23 November 2005
- Version 1.1.0 released 31 August 2006.
  - This version is frozen for use by the individual VxOs, but
  - development is continuing within the SPASE working group
- Resource Types: Catalogue
                         Display Data
                         Numerical Data
                         Granule
                         Instrument
                         Observatory
                         Person
                         Registry
                         Repository
                         Service

- Version 1.2.0 which will contain extensions based on community needs (probable release - mid 2007).

Future Users

The following data centres plan to provide SPASE descriptions of their holdings :

- All Virtual Observatories supported by NASA :
  - VHO (Heliospheric)
  - ViRBO (Radiation Belt)
  - VITMO (Ionosphere, Thermosphere and Mesosphere)
  - VMO/G (Magnetospheres)
  - VMO/U (Magnetospheres)
  - VSPO (Space and Solar Physics)
  - VSO (Solar Physics)
- ESA Cluster Active Archive (CAA)
- Centre de Données de la Physique des Plasmas (CDPP)

SPASE is not free !

- Individual data centres and repositories must have their own internal data model and dictionary, to exercise control over their own system. For example, dates of upgrades depend upon :
  - operating system upgrades,
  - major updates to support (*e.g.*, database or graphical) software, and
  - requirements to support launch and commissioning campaigns, special scientific workshops or similar projects.
- Therefore to be SPASE-compatible a data repository must convert as many as possible of its internal data descriptions to the current version of the SPASE model and dictionary. This is :
  - Relatively easy if the data model resembles the SPASE model (essentially translation of some keywords), but
  - Much more difficult if the data model is different ; in addition to keywords, at some level the metadata syntax must be understood.
- The conversion need not be perfect, but information not converted cannot be used by SPASE.
- Virtual Observatories which are deemed sufficiently SPASE-compliant may use the logo " SPASE inside ".

European Considerations

- There are currently three operational European data activities (in order of conception) :
  - CDPP (Centre de Données de la Physique des Plasmas), a national centre for archiving space plasma data which opened in October 1999
  - EGSO (European Grid of Solar Observations), a virtual observatory led by UCL-MSSL and established under the Information Systems Technology Priority of the EC FP5 (which has now ended).
  - CAA (Cluster Active Archive), opened in February 2006 and funded by ESA to make publicly available data of the highest possible quality from the four-spacecraft Cluster mission.
- How do these three entities fit into the scheme presented previously ?
  - EGSO is a virtual observatory, and should participate as such.
  - CDPP and CAA are magnetospheric data centres, which are respectively multi-mission and project-oriented. How do they fit in ?
- Is some European coordination desirable, or will European data integration be managed from NSSDC ?

Virtual Observatories and Space Weather

- The requirements for Virtual Observatory are primarily driven by the research community, whose priorities are :
  - location of data, generally long after the event
  - being able to exploit it without undue software development

- The requirements for an operational space weather service are different. The primary requirement is rapid access to the required data, which implies that :
  - it's location is known, and
  - processing software already exists.

- Despite this difference, Virtual Observatory services will be increasingly used for the research activities associated with Space Weather.

- Use of standard formats and data descriptions would probably accelerate (TBC) the integration of new data sources or services into an operational service.

Future Plans

- SPASE
  - Use the SPASE V1.1 data description to describe as many data sets as possible
  - Test the functionality and incorporate necessary modifications into V1.2
  - Improve the description at the physical parameter level to a level adequate to support API

- Europe
  - At the SPASE meeting in Toulouse in June 2006 one day was devoted to European activity
  - The presentations can be found at the URL
    http://www.cesr.fr/~harvey/SPASE_060608.htm
  - Many organisations are clearly motivated but working rather independently.
  - How can permanent European collaboration be established ?

End of the presentation

Thanks

Submitted Abstract

- As soon as the first data became available on-line over the Internet, it was obvious that different sites holding related datasets should appear to the end user as a single data system, even if the data itself is stored at multiple locations. In the realm of Space Plasmas, the international consortium SPASE (Space Physics Archive Search and Extract) was created in 2003 by several data centre. Simultaneously, several Virtual Observatories were being developed in the USA and Europe. Since 2005 US participation in SPASE has been supported by NASA, and early in 2006 NASA launched five new Virtual Observatories to cater for different aspects of Solar System plasma science. This presentation will outline the current status of this effort, the opportunities it offers, its specificities with respect to the Astronomic Virtual Observatory, and the possibilities it offers for space weather.