Solar and Space Physics Data in the EarthCube

Michael Wiltberger and L. Sitongia
National Center for Atmospheric Research

R. Weigel
George Mason University

Summary

The solar and space physics community produces a broad range of observational and simulation data. In this brief white paper we outline a set of requirements that the EarthCube data environment must support in order for it to be useful for our field. Key among these requirements are awareness of our data types, an easily utilized analysis interface and the ability to deal with massive datasets. Many of these requirements are not necessarily limited to the solar and space physics community, but applied broadly to the geosceinces.

Background

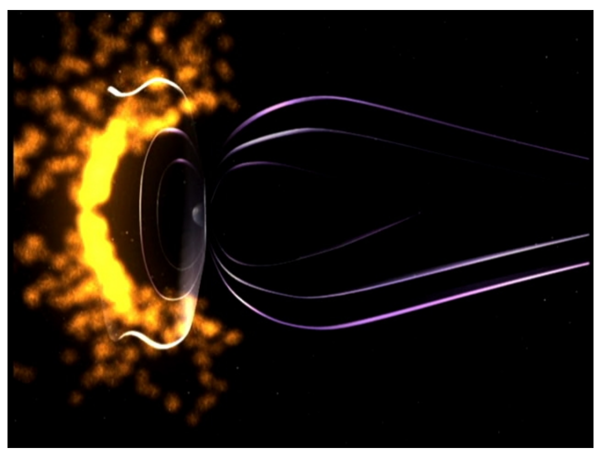

Figure 1 - Artist rendering of solar wind interaction with the Earth's magnetic field leading to the formation of the magnetosphere. Extracted from a NASA animation.

Solar and space physics are an important part of the geoscience field. The Sun is a major source of energy for the terrestrial system and is highly variable on timescales ranging from minutes to decades. Key components of the solar variability come from variations in solar radiative output and the state of the ionized gas, the magnetized plasma known as the solar wind that is constantly flowing from the Sun filling interplanetary space. When this plasma and the magnetic field imbedded in it reaches the geospace it interacts with the Earth's magnetic field forming the magnetosphere as illustrated in Figure 1. Geospace is populated with plasma that originates from either the solar wind or from the ionized upper levels of the Earth's atmosphere. This plasma supports coupling between the magnetosphere and ionosphere, a widely know reflection of this interaction is the aurora borealis or northern lights. The impact of this dynamic system extends beyond a beautiful light show and includes the field of research known as space weather. Impacts of space weather include disruption of radio communications, reduced accuracy of GPS navigation system, and damage or loss of satellites. The solar magnetic field also modulates the amount of cosmic rays striking the Earth's atmosphere, which alter ionization rates, affecting the abundance of aerosols that serve as condensation nuclei for cloud formation. Variability in solar radiative output causes changes in the chemistry of the Earth's upper atmosphere, particularly for solar UV wavelengths, which alter the production and loss rates of ozone that effect stratospheric heating.

A vast array of techniques are used to observe geospace both in space and from the ground. The Sun is routinely observed from the ground at several observatories supported by the NSF and other agencies. Numerous satellite missions, supported by NASA and other space agencies, have been launched to measure solar output at many different wavelengths. Satellite missions with plasma as well as plasma and magnetic field instruments fly throughout geospace and in the solar wind upstream providing in-situ measurements. Other satellite missions have used imaging techniques ranging from Visible and UV cameras to Energetic Neutral Atom imagers. On the ground the current trend is to uses spatially distributed arrays of instruments. Among these are networks of magnetometers that measure perturbations to the Earth's magnetic field caused by magnetosphere-ionosphere coupling, GPS stations that allow for determination of the electron density in the ionosphere, HF radars that use reflections from irregularities to diagnose the motion of the ionosphere, and all-sky camera networks that provide images of auroral emissions. The space physics data environment must be flexible enough to deal with all these different data types and sources.

Over the past decade the space physics community has invested significant resources to develop numerical models for the individual regions in the solar terrestrial system. More recently, work has been conducted to couple these models together to produce an end-to-end simulation of the entire system [*Luhmann et al.*, 2004]. These tools are then used to probe our physical understanding complex and highly coupled system. They are also beginning to make their way into the national infrastructure for space weather forecasting. Our community has also made these tools available to users beyond the developers through the Community Coordinated Modeling Center sponsored by NASA, NSF, and the DOD. Since geospace is so vast model, results will continue to be a key component of space physics data environment.

User Requirements

The users of space physics data in the EarthCube will include providers, both from observations and models, and consumers that will use the data to advance their scientific studies. In this section we highlight some of the requirements that an ideal data interworkability environment will provide to the space physics community. It is important to note that aspects of these requirements point toward universal requirements for all geoscience endeavors. Many of these requirements are now partially fulfilled, using many implementations, in an ad-hoc way; ideally the EarthCube initiative would support or foster cross-cutting geospace/space science integration activities that allow broad adoption of only a few implementations for each requirement.

Awareness of Space Physics Data

The data produced by space physics field covers numerous types of observations. The EarthCube data environment must be able to handle all these data types. Adoption and promotion of the Space Physics Archive Search And Extract (SPASE) metadata standards will be beneficial [*Thieman et al.*, 2010]. The system will also need to provide for the integration of data structures at the database level in order to provide access to the data required by the community and to minimize the demands on data providers. Implicit in this requirement is ability to support the announcement and attribution of datasets in the scientific literature.

Analysis Interface

An essential aspect of any data environment is the ease at which users at all levels of expertise can use the analysis interface. A beginning user will need to have an interface that presents them with a limited set of options to acquire the display the data. A more advanced user will want more sophisticated analysis and visualization tools, ones capable of conducting calculations not envisioned by the developers. The will have to provide an API for formulating and executing these operations. The most sophisticated users will need to adjust the API at the lowest level to allow for efficient and effective implementation of the tools needed by the community. We view successful implementation of this interface as an absolutely essential requirement to fulfill, if adoption of the data environment by the broad geospace community is expected.

Integration of Model and Observational Data

In our opinion the data environment, including the analysis interface, must be able to integrate data from both models and observations. It should be aware of the data types and grids used by simulations and observations and have the capability to conduct transformations needed to compare different outputs. A user of the environment should be able to load model results into the analysis interface and have it automatically query the environment to find all known observations and then be presented with the opportunity to conduct these comparisons. Support of the computation of standard error metrics would be a useful part of this system.

Community Analysis Tools

Even the best of developers will not able to envision all of the data analysis operations that users will need to conduct. The extensible nature of the analysis interface is a partial solution to this problem. Another part is support for community based analysis tools. This will allow beginning users not to have repeat work that has already been done while providing intermediate users with a forum for releasing their tools to the community in addition to the scientific publications that results from using the data environment.

Large Dataset Support

As computing power increases the size of datasets produced by simulations is outpacing the typical storage environment making the common paradigm of compute, move and analyze untenable. This is also true for observational datasets, for example NASA's Solar Dynamics Observatory mission is producing 1.5TB of data every day. Therefore it is essential that data environment support remote execution of analysis products in order to avoid overwhelming the network. On the modeling side this includes the need for remote rendering of large 3D data products.

Agency Neutrality

Observational data for the space physics community is produced by large number of agencies within the US government and throughout the world. The typical user of the data environment should not have to know where the data are stored or which agency produced it in order to acquire access. However, the user should be provided with sufficient information about the dataset to be assured of its quality and to provide attribution in scientific publications.

Intelligent Processing Pipelines

Processing of data often occurs in silos, preparing for assimilation and analysis by tools narrowly designed. Integrated analysis tools require that early data processing tasks be aware of other data sets, applications and analyses. Processing pipelines could assert requirements on other processing pipelines outside of their institution, so that the resulting data have parameters and keywords which will be directly useable by the integrated analysis tool, which could itself be the source of these assertions placed upon the processing pipelines. This is a process that is often done in an ad hoc fashion after a researcher recognizes the value of a new dataset. This requirement underscores the need to plan ahead of time for those eventualities.

Cross-cutting services

Over the years, many community-specific solutions have been developed that have the potential for use that spans both space and earth sciences. Integration and generalization of existing tools should be encouraged and emphasized. One paradigm for collaborative tool development is the Solarsoft software development project that is a library of software routines to support solar data analysis and spacecraft operations [*Freeland et al.*, 2000]. The library supports multiple space-and ground-based missions and is used throughout the world. Users can submit new software tools or update existing ones that have greatly expanded the scientific analysis capabilities of the library.

Education and Outreach

The old adage 'if you build it they will come' is really not true. They will only come if you advertise it. A key to broad adoption of the data environment will be efforts made to educate the scientific community about its capabilities and limitations. Teaching the next generation of scientific researchers about these tools will allow them to spend less time processing data and more time working to advance their science goals. Furthermore, an easily accessible data environment will provide a useful tool for promotion of science to the public and to teachers and students at the K through 12 levels.

Conclusions

The space physics community will likely benefit greatly from the implementation of the EarthCube data environment assuming it meets the requirements outlined above. In this ideal environment, we envision a new researcher being able to, for example, conduct a high-resolution simulation of the interaction of the solar wind with the magnetosphere-ionosphere system. This run will be included in the database of events and using the metadata standards and integration across agencies they will be provided with a menu of possible verification comparisons that they may wish to have conducted. The user will be able to select from a broad range of advance analysis tools developed by other users of the system to further refine their investigations. In the final stages of their research, they will develop tools to complete their analysis and publish their results.

References

Freeland, S., R. Bentley, and P. Murdin (2000), SolarSoft, in *Encyclopedia of Astronomy and Astrophysics*, P. Murdinedited by.

Luhmann, J. G., S. C. Solomon, J. A. Linker, J. G. Lyon, Z. Mikic, D. Odstrcil, W. Wang, and M. Wiltberger (2004), Coupled model simulation of a Sun-to-Earth space weather event, *Journal of Atmospheric and Solar-Terrestrial Physics*, *66*(1), 1243–1256, doi:10.1016/j.jastp.2004.04.005.

Thieman, J. ~., T. ~. King, D. Roberts, Spase Consortium (2010), SPASE 2010 - Providing Access to the Heliophysics Data Environment, *AGU Fall Meeting Abstracts*, A1862.