EOS
EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 90 NUMBER 47
24 NOVEMBER 2009
PAGES 441–452

Using Virtual Observatories for Heliophysics Research

PAGES 441–442

Scientific satellites, balloons, ground-based instruments, and other observational platforms are producing rich streams of data about the Earth and space. Ensuring widespread access to such data has led to the development of a new type of observatory: the virtual observatory. Existing only in cyberspace, virtual observatories are Web-based interfaces that point users to online data repositories. More important, they allow users not only to access and view multiple sources of information at the same time but also to cross-compare data to build new insights.

A number of new virtual observatories have opened their doors to users interested in researching heliophysics, setting the stage for new discoveries about the nature of the Sun-Earth connection made possible by the variety and quantity of data and tools available using a virtual observatory. These observatories continue to evolve as new high-level interfaces and services are developed to meet modern research demands.

What Is a Virtual Observatory?

Virtual observatories seek to enable an unprecedented level of integration and access to data and model resources. The vision for achieving this integration is through discipline-specific virtual observatories that use techniques derived from the new field of informatics and data sciences [*Baker et al.*, 2008] for intercommunication. In the continuum of the data life cycle, the virtual observatory is the front end to the extended efforts that are required to fund, acquire, calibrate, and synthesize data. In the same way that office management tools such as e-mail, word processors, and photo galleries are becoming increasingly Web-based, thereby opening up new avenues of access, communication, and collaboration, the virtual observatory seeks to provide enhanced access, communication, and collaboration for science research and analysis.

A virtual observatory is a service, implemented as software applications, for (1) providing uniform and easy access to discipline-specific data from many different data providers and (2) providing services for both data access and science analysis. It is also responsible for developing the different types of metadata required for facilitating these services. The types of metadata include those for use by scientists and those for use by software.

Virtual Observatories and Heliophysics: A Retrospective

For heliophysics, virtual observatories enable the large-scale cross-disciplinary research required for study and prediction of the Sun-to-Earth plasma environment. The role of these virtual observatories is separate and distinct from those of resource providers and data producers. The relationship between the virtual observatory and the greater heliophysics data environment is shown in Figure 1.

Virtual observatories form an integral part of NASA's Heliophysics Data Environment (HPDE; http://hpde.gsfc.nasa.gov/) and were motivated by the needs outlined in the "Report and Recommendations of the Living With a Star Science Data System Planning Team" (http://hpde.gsfc.nasa.gov/LWS_Data_System_Final.html). In fact, this need for Web-based communities is so strong that within heliophysics, several subdiscipline-specific virtual observatories (VxOs, where "x" represents a subdiscipline such as "solar") have sprung up or are being built as part of HPDE's efforts.

The Virtual Solar Observatory (VSO; http://virtualsolar.org/) was the first heliophysics virtual observatory, established in 2001. It links to data from ground- and space-based instruments that observe the Sun, and many of the concepts from its design have been integrated into recommendations for future VxO development, as reflected in NASA's newly released Heliophysics Science Data Management Policy (http://hpde.gsfc.nasa.gov/HPDP.html).

Several other VxOs followed. The Virtual Space Physics Observatory (VSPO; http://vspo.gsfc.nasa.gov/) provides access to a comprehensive list of space physics data products and data repositories; in addition, it provides general search facilities and retrieval of data sets from the Space Physics Data Facility (http://spdf.gsfc.nasa.gov/) into which it was recently integrated. The Virtual Ionosphere/Thermosphere Observatory (VITMO; http://vitmo.jhuapl.edu) provides access to a large selection of data products and includes advanced search capabilities. The Virtual Radiation Belt Observatory (ViRBO; http://virbo.org/) has many of the standard VxO search and retrieval capabilities and has made accessible a number of previously offline data sets. ViRBO is working with the instrument and data teams of the upcoming Radiation Belt Storm Probe Mission (RBSP) to educate them on the many new VxO tools and services that they can leverage. The Virtual Magnetosphere Observatory (VMO; http://vmo.nasa.gov) is developing a comprehensive inventory of magnetosphere data products and an advanced search and query interface intended for magnetosphere science needs.

A more comprehensive list of virtual observatories is available at http://hpde.gsfc.nasa.gov/.

Metadata in VxOs

The initial phase of the virtual observatory effort has involved the development of metadata that document the holdings of data providers along with the software infrastructure that enables connections among VxOs. Of course, the ultimate goal is to develop research services, and in a short amount of time, scientists will likely be able to report on the science-level activities enabled by virtual observatories.

There has been widespread recognition of the value of and need for a common metadata model/data dictionary that encompasses all of heliophysics. The development of a common metadata model for interoperability at the semantic level is a necessary and important step required for achieving cross-VxO interoperability at the data level. The very active Space Physics Archive Search and Extract (SPASE) metadata model group (http://spase-group.org/), which includes participants from every heliophysics subdiscipline, has undertaken this task. The work of the group has led to mature

By R. S. Weigel, D. N. Baker, D. A. Roberts, and T. King

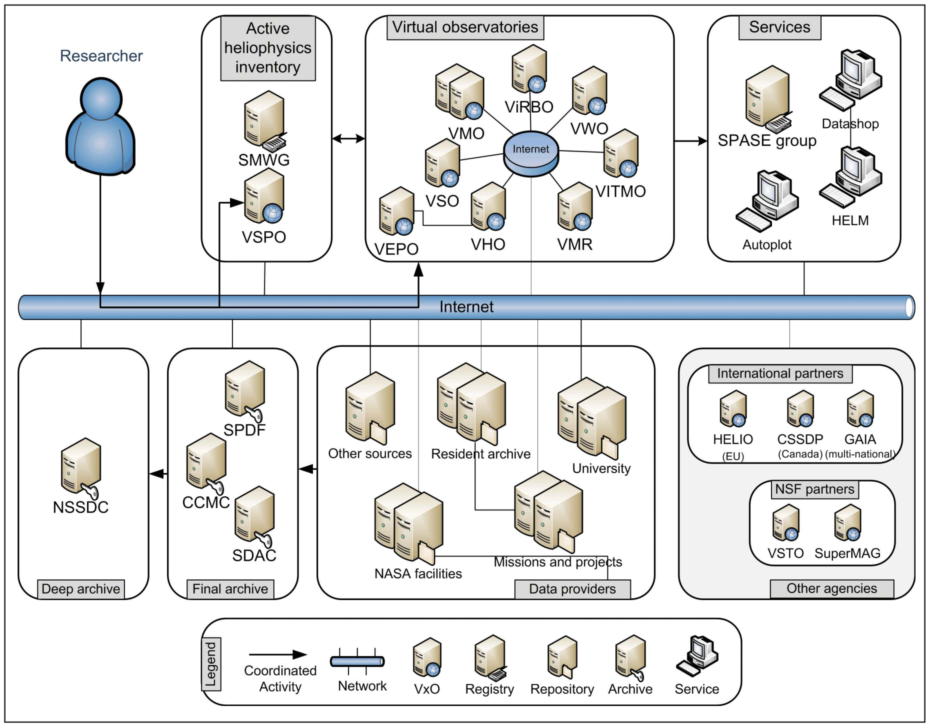

Fig. 1. A researcher has many routes to data and services. One route is to access data providers directly through the Internet by using general Internet search engines. Or, using the Virtual Space Physics Observatory (VSPO), a researcher can do a coarse-grained search on the inventory of available heliophysics data. This inventory is created and maintained by the VxOs. Alternatively, the researcher may directly visit the VxO associated with his or her specialty to perform fine-grained and discipline-specific searches and have access to discipline-specific services and community resources. The virtual observatories include the Virtual Energetic Particle Observatory (VEPO), the Virtual Solar Observatory (VSO), the Virtual Magnetosphere Observatory (VMO), the Virtual Radiation Belt Observatory (ViRBO), the Virtual Wave Observatory (VWO), the Virtual Ionosphere/Thermosphere Observatory (VITMO), and the Virtual Model Repository (VMR). The services include the Heliospheric Event List Manager (HELM), the Space Physics Archive Search and Extract (SPASE) discussion group, and two other software projects: Autoplot and Datashop. The deep archive is the National Space Science Data Center (NSSDC), while the final archives include the Space Physics Data Facility (SPDF), the Community Coordinated Modeling Center (CCMC), and the Solar Data Analysis Center (SDAC). International partners include the Heliospheric Integrated Observatory (HELIO), the Canadian Space Science Data Portal (CSSDP), and Global Auroral Imaging Access (GAIA). VxOs sponsored by the National Science Foundation (NSF) include SuperMAG and the Virtual Solar Terrestrial Observatory (VSTO).

recommendations adopted by most VxOs on common nomenclature.

The VxOs each have their own way of internally handling and managing metadata. One virtual observatory, ViRBO, is developing a metadata content management system with comment, rating, and wiki features (http://virbo.org/meta/) that has cross-VxO applicability and can be used as a replacement for ad hoc management systems. The latest release of the SPASE data model includes a method to describe and share annotations. With the advent of more data producers, including modelers who provide simulation data sets, certain features will become important as groups seek to post their metadata in a place where they get maximum exposure, and new users will want to be able to determine the quality of the data set by looking at the data set discussions and its associated ratings.

Another key success is the development of a version-controlled SPASE metadata repository for metadata that are of interest to multiple virtual observatories (such as satellite and instrument metadata and contact information for people associated with data sets). This repository now has more than eight active contributors from four virtual observatories. The result has been an overall improvement in metadata quality and accuracy and general agreement on metadata usage standards. Many of the virtual observatories also maintain version-controlled repositories for their discipline-specific metadata.

In addition, metadata ingestion services will become a necessity. These services will allow users to easily contribute metadata (and possibly small data sets) and have them immediately available to the heliophysics community.

The Need for Advanced Searches

Several virtual observatories have advanced search capabilities, which include the ability to do complex queries that restrict results based on conditions that include date ranges, instrument type, position of the instrument in space, and many other elements that are a part of the SPASE vocabulary. However, early feedback from some longtime community members has been that for them, there just are not enough data for advanced searches to be a "must-have" feature; in many cases, browsing a directory tree with a listing of satellites and instruments is sufficient because they know what they are looking for. For this reason, the virtual observatories will sometimes provide access to data in a more traditional way using the directory tree list metaphor.

Soon science-grade data will be available from low-cost small satellites [*Moretto and Robinson*, 2008; *Baker and Worden*, 2008]. In this case the number of instruments and data sets will grow significantly. Thus, advanced search features will become a necessity, and they are already a convenience given the hundreds of new products that have been registered at VxOs over the past 3 years. The need for advanced searches has also recently increased as virtual observatories have gained access to a number of new data sets and are in the process of releasing many new services. It is clear that there will need to be a change in user behavior from browsing categorized lists to scanning search results that tap into comprehensive databases of data products that are updated and maintained by the entire heliophysics community.

New Solutions and New Needs

One of the less prominent, but key, successes in the virtual observatory development effort has been in software sharing and the leveraging of existing software code bases. Scientists tend to try to develop or work through problems on their own, especially on software problems where the solution is obvious but just requires time. This is changing, as evidenced by an effort to provide developer tools and documentation for code bases that can be shared (http://teamvxo.org/). Thus, those who are now developing products for new missions will likely have a much different experience from those who developed data products for previous missions, as many new visualization, online data sharing, and versioning tools will be available for use without requiring in-house development. Indeed, some of the new visualizations commissioned by mission teams will be immediately and easily available to the VxO community because (1) they will use and build upon tools that are developed and used by VxOs and (2) some of the instrument team developers and scientists are also part of the virtual observatory community and have a vested interest in promoting software sharing.

But hand in hand with new solutions come new demands for services. Many of these are technically outside of the virtual observatory purview, but the VxOs have stepped in to provide these services because doing so will result in an overall enhancement of the data environment:

• *New data:* A number of VxOs are involved in non-VxO tasks: holding and storing data. The discipline-specific nature of the VxO is partly the reason; the VxO leads have close contacts with scientists in the field and interact with them at meetings. Another reason is that some data do not have permanent homes after a project has ended. The VxOs have stepped up to provide one.

• *New products:* Scientists who have data sets for a paper want to make them available to their community, and they realize that their discipline's VxO would be a natural place for storage. In the past a scientist who developed a specialized data product as a part of a proposal would post it on a personal home page. Besides the obvious problems with longevity of personal pages, visibility and searchability of these data products are reduced. VxOs ease this problem.

• *New code:* A number of scientists realize that the best way to introduce and promote software codes is through their discipline's virtual observatory. Rather than storing it on a generic software repository, they make the primary storage point the virtual observatory.

The eventual goal of the virtual observatory is to allow high-level science analysis. Now that the infrastructure is in place, the next step is to bring in the users. In the next few years, VxO users are expected to gradually change their familiar methods of data access and exchange to methods that leverage the new technologies and services available through the VxO.

What Can Be Expected for the Future?

Consider how a photo gallery was displayed in the early days of the Internet on a personal Web page. For the early Internet adopters, it was straightforward to write html and upload files to a Web page. New photo-sharing sites made this process even easier and allowed nontechnical experts to do it with ease. The major benefit of such sites is the aggregation of community resources that allowed for a single application programming interface (API) and user interface for searching, tagging, sharing, and commenting on photos. Although standard Web searches on images still work, domain-specific image services provide an even higher level of utility. The same applications hold true for virtual observatories, and heliophysics community members are now becoming more aware of how to use the added-value services developed by VxOs.

The most important task for virtual observatories is to integrate new services that meet a discipline-specific science need. Further, there is a growing understanding in the VxO developer community of the importance of interoperability at the data level. A number of existing APIs for data, including the Open-Source Project for a Network Data Access Protocol (OPeNDAP; http://opendap.org/), appear to satisfy many of these needs, and several heliophysics VxOs plan on providing their data using this protocol, which has extensive support and use in the climate and atmospheric sciences community.

The long-term success of the virtual observatory depends on the close interaction with the science community. The separation of the virtual observatories into subdomains will help foster that interaction and at the same time make data accessible from other subdomains through an interface at a VxO relevant to another subdomain. Through these links, the collection of VxOs will form one conglomerate, the "Heliophysics Great Observatory." With the metadata and software infrastructure maturing, the next phases involve (1) the iterative process of refinement and revision based on user needs and expectations, (2) the continual addition of new services and data resources, and (3) the promotion of communication and data sharing through virtual observatories and within subspecialty communities of the HPDE.

References

Baker, D. N., and S. P. Worden (2008), The large benefits of small-satellite missions, *Eos Trans. AGU*, *89*(33), 301, doi:10.1029/2008EO330001.

Baker, D. N., C. E. Barton, W. K. Peterson, and P. Fox (2008), Informatics and the 2007–2008 Electronic Geophysical Year, *Eos Trans. AGU*, *89*(48), 485, doi:10.1029/2008EO480001.

Moretto, T., and R. M. Robinson (2008), Small satellites for space weather research, *Space Weather*, *6*, S05007, doi:10.1029/2008SW000392.

Author Information

Robert S. Weigel, Department of Computational and Data Sciences, George Mason University, Fairfax, Va.; E-mail: rweigel@gmu.edu; Daniel N. Baker, Laboratory for Atmospheric and Space Physics, University of Colorado at Boulder; D. Aaron Roberts, Heliophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, Md.; and Todd King, Institute of Geophysics and Planetary Physics, University of California, Los Angeles

Field Geophysics Class at Volcán Tungurahua, Ecuador

PAGE 442

Ecuador's erupting Volcán Tungurahua was the recent site of a 3-week graduate-level geophysical course on volcanoes, hosted by Ecuador's Instituto Geofisico of the Escuela Politecnica Nacional (IG-EPN) and the Department of Earth Science at the New Mexico Institute of Mining and Technology (NMT). Sixteen students from 12 universities and four countries participated in the intensive course, which entailed broadband seismometer and infrasound sensor deployment followed by subsequent data processing, analysis, interpretation, and result synthesis. Hardware for the course was provided by the Incorporated Research Institutes for Seismology (IRIS) through the Program for Array Seismic Studies of the Continental Lithosphere (PASSCAL) as well as the IG-EPN and NMT geophysics programs.

Since the start of its most recent eruptive period (in 1999), Tungurahua has proved itself a reliable source of both seismicity and infrasound radiating from its typically open vent. As such, Tungurahua provides the ultimate outdoor teaching laboratory where students can deploy instruments for just a few days and then collect earthquake and explosion data. Tungurahua's activity in June 2009 did not disappoint class participants: Frequent earthquakes included long-period and volcano tectonic events, various types of tremor events, and explosion earthquakes manifested by booming "cannon-shot" blasts. Some of the explosion shock waves were recorded 10 kilometers from the vent with excess pressure amplitudes greater than 50 pascals in the infrasonic band. Had these intense sounds been audible, their sound pressure levels at 10 kilometers would have been in excess of about 130 decibels!

The manifestation of the seismic tremor events was equally dramatic. One of these tremor events was witnessed up close and personal by the class as they approached a field site just in time to watch a torrential lahar of rain-mobilized ash flow down one of Tungurahua's drainages. Small explosive (Vulcanian) and lava fountain (Strombolian) eruptions were common daily and were especially satisfying to watch while knowing that the corresponding elastic wave radiation was being recorded.

After deployment of a seven-station seismic array and a 10-component infrasonic network (see Figure 1), the class set out on a whirlwind volcano road trip while waiting for earthquake data to accumulate. With guest instruction from Pete Hall, Patty Mothes, and Patricio Ramon (IG-EPN) and Claude Robin (Institut de Recherche pour le Développement (IRD), France) the class toured the ash flow deposits of Quilotoa caldera, visited Cotopaxi National Park and Chacana caldera deposits, climbed to the rim of Guagua Pichinchas (which erupted in 1999–2000), and toured the currently erupting Reventador volcano in the eastern jungle lowlands. Stops at the Tungurahua and Quito volcano observatories and examination of devastation from recent eruptions at Tungurahua, Cotopaxi, and Reventador added