The Virtual Solar-Terrestrial Observatory (VSTO)  
December 2005

Project Description
The goal of this project is to develop at the National Center for Atmospheric Research (NCAR), based within the High Altitude Observatory (HAO) and in collaboration with the Scientific Computing Division (SCD), a prototype for a Virtual Solar-Terrestrial Observatory (VSTO).

This project addresses key science areas such as solar-terrestrial physics datasets, the highly interdisciplinary Center for Integrated Space-Weather Modeling (CISM) model intercomparison, providing a framework for collaboration and a basis for building and distributing advanced data assimilation tools for the solar-terrestrial physics community. This project will directly address key needs in Cyberinfrastructure (CI) such as software tools and services, interdisciplinary data integration, representation, metadata, documentation, quality control and user community building.

Definition
The prototype Virtual Solar-Terrestrial Observatory (VSTO) is a distributed, scalable education and research environment for searching, integrating, and analyzing observational, experimental and model databases in fields of solar, solar-terrestrial and space physics (hereafter referred to as SSTSP).

VSTO comprises a framework which provides virtual access to specific SSTSP data, model, tool and material archives containing items from a variety of space- and ground-based instruments and experiments, as well as individual and community modeling and software efforts bridging research and educational use. The prototype will be a fully functional system directly addressing the immediate and substantial needs within the SSTSP community, allowing science projects to advance more rapidly. E.g. in solar coronal physics there is a need to cohesively assemble multi-wavelength images of the dynamic solar upper atmosphere.

Space weather model inter-comparisons, and Assimilative Mapping of Ionospheric Electrodynamic results need to be distributed to their communities. Solar activity indicators and solar total and spectral irradiance data and models need to be made available to terrestrial atmospheric researchers for use in their studies.

What problem does VSTO address?
In discussions with data providers and users, the needs are clear: ``Fast access to `portable' data, in a way that works with the tools they have; information must be easy to access, retrieve and work with."

Too often users (and data providers) have to deal with the organizational structure of the data sets which varies significantly --- data may be stored at one site in a small number of large files while similar data may be stored at another site in a large number of relatively smaller files.

There is an equally large problem with the range of metadata descriptions for the data. Users and providers are still frustrated with the use of data and knowing or specifying its heritage. Users often only want subsets of the data and struggle with getting it efficiently. In summary one user expresses it as: ``(Please) solve the interface problem.'' VSTO addresses this specific problem.
Since there are an increasing number of discipline-specific virtual observatories either operating or under development, we propose to scope this interdisciplinary virtual observatory as follows: the research fields of solar, solar-terrestrial and space physics are substantially interdisciplinary already, the demand to provide an interoperable information exchange infrastructure is increasing, the practical needs of data providers and data consumers are clear, and this project provides an opportunity at addressing a fundamental computer science question of knowledge representation across disciplines.

It is our expectation that progress in building a VSTO will shed light on how virtual observatory principles can be applied to a variety of other scientific disciplines and, in fact, bridge disciplines.

It is also our premise that a substantially positive impact will result for the SSTSP community through this prototype because there is both a significant amount of scaffolding already in existence and the next steps and tools to address the interdisciplinary problem are now clear and available (and are presented within this proposal).

The specific tasks proposed are divided into two areas: cyberinfrastructure development, and the demonstration of this infrastructure in the current science data and modeling environment of HAO.

Technology capability - ontology development and evolution
Knowledge about application tools: how they read data, how they store, and what the output data represent, are specific problems to address, especially when accessed through distributed computing resources. Problems quickly arise when trying to understand interactions among complex or interdisciplinary systems; something that exists in SSTSP for both education and scientific research processes. Too often, the specific storage formats and protocols become an additional barrier in opening the information up to another area. Even today, the products educators and scientists seek are based in complex structures and/or contain or rely upon large volumes of reference materials, (often-undocumented) computer code, visual representations, and/or numeric data. Existing standards such as Open Knowledge Base Connectivity, and the Knowledge Interchange Format and existing markup standards such as OWL and RDF-DAML will be leveraged as a basis to record content such as application interfaces and data outputs.

Thus, a key technology development will be a set of linked, i.e. interdisciplinary-enabled, ontologies arising out of VSTO-specific information along with integrating and merging existing semantic metadata.

This process requires tools that support broad ranges of users in (1) merging of ontological terms from varied sources, (2) diagnosis of coverage and correctness of ontologies, and (3) maintaining ontologies over time.

Contact us:
Peter Fox (PI) pfox@ucar.edu
http://vsto.hao.ucar.edu
http://www.vsto.org/
The Virtual Solar-Terrestrial Observatory (VSTO)  
December 2005