

Heliophysics Data Environment Enhancements
Abstracts of selected proposals.
(NNH09ZDA001N-HDEE)

Below are the abstracts of proposals selected for funding for the Heliophysics Data Environment Enhancements program. Principal Investigator (PI) name, institution, and proposal title are also included. 18 proposals were received in response to this opportunity, and 11 were selected for funding.

Thomas Armstrong/Fundamental Technologies, LLC
HISCALE Fundamental Technologies Resident Archive Data Services Continuation

It is proposed here that Fundamental Technologies, LLC will organize, validate, and create the most time-efficient, cost-effective, access to the entirety of the Ulysses HISCALE data. Because the data occurs at 12 second cadence for approximately 19 years, a large number of individual files have been generated (upwards of 100,000 files in total). We will use some automated methods to assure the basic integrity of the files to be released via the Virtual Heliospheric Observatory/Virtual Energetic Particle Observatory accompanied by SPASE-compliant XML descriptions of each product type and individual file (granule). Since the data are released in fully calibrated form as species-identified, directionally distinguished, fluxes and energy calibrated events, it is important to verify the usability and reliability of these data for science purposes. This proposal will allow the continuation of our efforts to test our documentation, the VHO/VEPO data set, and ancillary data sources in which we use undergraduate physics majors to carry out a typical science investigation. The validation of the archives is achieved primarily by showing that the data are sufficient for successful science investigations by novice users from outside the HISCALE investigation.

Linton Floyd/Interferometrics Inc.
Improved Wavelengths for SUSIM High Resolution Solar Ultraviolet Irradiance Spectra

The objective of this project is to improve solar UV spectrameasured by the Solar Ultraviolet Spectral Irradiance Monitor(SUSIM) aboard the Upper Atmosphere Research Satellite (UARS)by improving the wavelength assignments. These spectralmeasurements were gathered at 0.15 nm resolution over a periodof nearly 14 years from 1991 to 2005. These wavelengthassignments can be improved in two ways. First, thewavelengths of important and recognizable solar features, theso-called "fiducial features", distributed throughout thesolar spectrum will be reassigned based on values derivedthrough examination and convolution to the SUSIM 0.15resolution of standard solar spectra having higher resolution.Interpolation of these fiducial assignments onto theindividual wavelength measurements will be conducted based onan improved understanding of the original hardware-basedwavelength assignments. Once the wavelengths have beenrecalculated and a consequent minimal adjustment to theirradiance values is

performed, the data and metadata will be packaged into formats acceptable to the VSO and other solar data repositories.

Corrected wavelength assignments are expected to aid in the understanding of the spectral dependence of solar irradiance variation. It is expected that the correspondence with other solar spectra such as those measured by SCHIAMACHY and SORCESIM will be improved. Establishing a correspondence with these twenty first century UV irradiance measurements will promote our understanding of the long term UV spectral irradiance record (especially at higher resolution) that currently spans more than three decades. Further improvements in the SUSIM high resolution wavelengths will refine the correspondence among those spectra. The advantage is that irradiance time series at individual wavelengths will more sensitively reflect the underlying solar activity because of the resulting reduction in noise. There are at least two ways in which solar irradiance science is advanced by these refined time series. First, better estimates of the irradiance variation over the solar activity cycle would be achieved through comparisons with solar indices (e.g. Mg II core-to-wing ratio and the Photometric Sunspot Index). Second, the 0.15 nm resolution is sufficiently high so that more granular understanding of the variation with respect to its solar atmospheric origins would be possible.

Umran Inan/Stanford University

Data Services Upgrade: Restoration of the DE-1 Plasma Wave Instrument Wideband Data Set

A one year program is proposed with the purpose of restoring and making available to the scientific community wideband plasma wave data from Dynamics Explorer 1 (DE-1) Plasma Wave Instrument (PWI). The only remaining copies of the wideband PWI data are currently stored at Stanford University on analog magnetic tapes of which approximately 1000 exist. We propose to digitize the raw data, create several levels of scientifically useful data products fully documented with appropriate meta-data, and make all products available to the broad scientific community through the Virtual Wave Observatory (VWO) and the Coordinated Data Analysis Web (CDAWeb). The DE-1 PWI wideband data has high scientific value for ongoing and timely studies of plasma waves, particularly those that contribute to the acceleration and loss of highly energetic radiation belt particles.

Homa Karimabadi/SciberQuest, Inc.

Value Added Services for VxOs: Automated Science Discovery Tools

We propose to develop an intelligent data analysis engine that extends the functionality of Virtual Observatories (VxOs) from data portal to science analysis resource. Heliophysics is a data centric field which relies heavily on the use of spacecraft data for further advances. The prevalent approach to analysis of spacecraft data is based on visual inspection of data. As a result, the vast majority of the collected data from various missions has gone unexplored. While the need for advanced algorithmic approach to data exploration and knowledge discovery is generally recognized by experimentalists, the adoption of such techniques (“data mining”) has been slow. This has been partly due to

the steep learning curve of some of the techniques and/or the requirement to have a working knowledge of statistics. Another factor is the existence of a plethora of data mining approaches, and it is often a daunting task for a scientist to determine the appropriate technique. Our goal has been to make such tools accessible to non-experts and remove it from gee-whiz domain to a practical tool that will become part of the standard arsenal of data analysis. To this end, we have developed an automated data mining technique called MineTool. Its first deployment to analysis of Cluster was very successful and it is rapidly gaining adoption among experimentalists. There are currently several ongoing projects involving MineTool in magnetospheric and solar physics. The goal of this proposal is to leverage our recent success and offer a complete solution for data analysis as value-added service to VxOs. The proposed tasks are: i) Development of automated data preparation algorithms, ii) Integration of data preparation and data mining algorithms with VMO/VHO, iii) Use of our group of beta testers, experimental space physicists, to obtain feedback on the ease of use and functionality of the service and incorporation of the feedback into the design process, and iv) A real life deployment of the software for creation of event lists by our collaborators. These innovations will significantly enhance the science return from NASA missions by providing data centers and individual researchers alike an unprecedented capability to mine vast quantities of data.

Justin Kasper/Smithsonian Astrophysical Observatory
Data Services Upgrade: Calibrated Wind Faraday Cup Reduced Distribution Functions

The objective of this Data Upgrade proposal is to create and distribute calibrated solar wind proton and alpha velocity reduced distribution functions (VDFs) based on observations by the Faraday Cup instrument on the Wind spacecraft. While almost every NASA spacecraft (e.g. IMP-8, Voyager, Wind, SOHO, ACE, STEREO) with in situ heliophysics science objectives has an instrument that measures the phase space density or VDF of solar wind ions, generally data users in the greater community only have access to either the raw observations or "bulk" solar wind parameters (velocity, density, and temperature) extracted from the raw measurements by the experiment team. The inability of many users to easily view solar wind VDFs directly is a shortcoming because it is often the case that the solar wind is not well-described by a simple, single density, temperature, or velocity. Helium and minor ions at times can be significantly hotter and faster than protons, and the VDF of a single species can be highly anisotropic (temperature anisotropy), possess heat fluxes, or even break up into multiple components (such as the proton core and halo). By producing a calibrated data set of the ion reduced distribution function along different angles, general users will be able to investigate how well the bulk parameters describe the observations for any set of Wind observations. This will permit new scientific investigations of topics such as non-Maxwellian solar wind kinetic physics, shock physics, the lunar wake, and magnetospheric phenomena.

James McFadden/University of California, Berkeley
FAST Data Archive: Data Services Continuation

We propose to provide a “data services continuation” by maintaining the FAST Resident Archive located in the Space Science Laboratory at U.C. Berkeley. The FAST Mission was designed to identify and understand the physical processes that comprise the origin and consequences of the Earth's aurora. Measurements of ions and electrons, composition, and DC and AC electric and magnetic fields were collected as FAST passed through the 'average' location of the aurora near the Earth's magnetic poles (~11 orbits per day). Data were collected at several different rates, with the highest time resolution data collected 2-3 orders of magnitude faster than has ever been achieved by any other magnetospheric mission. During its ~13 years of operation, FAST has collected ~7.5 Terabytes of data. The FAST data analysis and data access software/hardware were designed for speed in a time when computers were much slower, and where computer memory and disk space were in short supply. Limited resources only allowed software development for a single platform (SUN-Solaris), which supported real-time data displays for I&T and FAST Campaigns, and standard data analysis displays. The data set resides in an on-line archive (RAID) allowing instant access to the entire data set. We propose to maintain this Resident Data Archive for the next three years, continuing to provide high resolution FAST data to the entire scientific community. This Data Services Continuation is essential not only for data analysis continuity by a large number of both national and international scientists, but also to provide a stable transition to a long-term Final Archive. In a separate proposal led by Dr. Strangeway (UCLA), funding is requested under the “Added Value Service for Virtual Observatories” to construct a Final Archive of FAST legacy data products, which should remain useful for decades to come. This proposal's data services continuation is required to complete this Final Archive.

Jeff Morrill/Naval Research Laboratory
The Data Upgrade of Recently Digitized Skylab UV Spectral Data from the S082 Instrument

The goal of this proposed program is to provide an accessible electronic database of spectra of specific solar surface features that can then be used to generate historical estimates of full disk solar irradiance spectra for use in high fidelity, long term climate studies as well as studies of the energetics solar atmosphere both through observations of the solar UV spectrum. Current efforts to understand the roles of anthropogenic and natural forcing on climate variability are beginning to focus more closely on the impact of solar variability on changes in climate. To examine the sun's role effectively requires an understanding of long-term solar irradiance variability and how this variability impacts the complex atmosphere/ocean/land system, for example, the response of stratospheric ozone to solar UV irradiance variability and how this response couples into atmospheric dynamics. Studies of the solar corona and transition region use UV spectral observations to develop and test models of the solar atmosphere. The UV observation from the Skylab S082 instruments will provide a unique and comprehensive data set to address these problems.

Bodo Reinisch/University of Massachusetts Lowell
Data Services Continuation for Resident Archives of IMAGE/RPI Expert-derived Data

The capability of the Radio Plasma Imager (RPI) of NASA's IMAGE mission to actively sense distant magnetospheric plasmas at 50,000+ km distances away from the spacecraft has been capturing the attention of the Heliophysics research community for years. Much remains undiscovered in the RPI dataset: it poses many great difficulties for a casual space scientist to identify significant wave data features and associate them with the underlying physical phenomena. As a result, serving and providing understanding of the IMAGE RPI data resources to the space research community relies on the processing and documentation of the RPI data by a distributed team of experts whose cumulative knowledge is captured, archived, and made available online using custom interactive software tools from two existing RPI Intelligent Resident Archives (RIRA), one at the University of Massachusetts Lowell (UML) and the other at the NASA Goddard Space Flight Center (GSFC). We propose to formally establish a unified IMAGE/RPI Resident Archive by coordinating the two RIRAs for effective serving of the IMAGE/RPI instrument data and related resources.

Since December 2000, thirty seven experts from 9 organizations in the USA and Europe have been affiliated with the existing RIRAs that, in addition to the full resolution telemetry, orbital, and Geospace modeling data resources, manage a unique collection of higher level data products derived manually and automatically from the otherwise unfathomable two million RPI plasmagram and spectrogram records. The Lowell RIRA provides software tools and an Internet-accessible central database with read/write access to support expert data analysis (interactive derivation of secondary data products and image annotation). Online connection to Lowell RIRA allows expert users to submit their analysis results to the database immediately and seamlessly as they work with the data. Thus acquired expert knowledge serves not only an educational purpose, presenting examples of data interpretation to the community, but also as a valuable, growing catalog of analytical, phenomenological information derived from the RPI data, searchable by content. The unique RIRA collection includes 3568 manually-annotated plasmagram images, in which remote RPI signal reflections are extracted for their subsequent inversion to produce magnetospheric electron-density profiles mainly along the magnetic field line through the image satellite. The Goddard RIRA serves as the WWW publishing portal that provides single-stop, easy and convenient public access to acquired expert interpretation data, display products, and documentation.

This proposal calls for continuing service of the Lowell and Goddard RIRAs, as well as a “rescue for posterity” effort to search and load additional, currently disorganized nuggets of the expert-derived RPI data interpretations available at UML and GSFC. We will appropriately improve metadata, documentation, and Web interfaces so as to facilitate visibility, understandability, and utility of the RPI products. We will also modify the existing data search facilities of the Lowell RIRA to support Space Physics Archive Search and Extract (SPASE) model interface, which will effectively allow servicing

queries from the NASA Virtual Wave Observatory (VWO) and thus accomplish much anticipated integration of the RIRA into the VWO domain.

Our ultimate goal can be summarized as follows. We propose that the Lowell and Goddard RPI Intelligent Resident Archive (RIRA) (1) continue acquisition and preservation of the unique expert knowledge derived from the RPI raw data, (2) load previously derived products for storage, search, and retrieval using RIRA services, and (3) change accompanying metadata and user interfaces (e.g., via the NASA VWO) so as to facilitate visibility, understandability, and utility of the higher order RPI science data that will enable new Heliophysics research projects.

Jack Scudder/University of Iowa

A Resident Archive (RA) for the GGS-Polar-Hydra-Data: "Data Services Continuation"

This proposal seeks funding to establish, and maintain a Resident Archive at the Principal Investigator's institution, University of Iowa, for the purpose of serving GGS-Polar-Hydra data acquired between February 24, 1996 and April 28, 2008. Because of its compactness the data are stored as in telemetry, with fast computer programs controlling the digest of selected intervals into the desired geophysical variables that are usually fluid moments, or spectrograms which will be supplied by the proposed RA facility. The RA will always provide standard spectrograms for survey purposes with photoelectrons removed and energies shifted to be assays of the local plasma well away from the spacecraft. They will also contain overlay traces of the evolution of the mean energy of the ions and electrons - without characterization of its partition between temperature or flow energy. In addition the RA will provide legacy data in the form of the ion and electron density, flow velocity, parallel and perpendicular pressures and heat flux as determined by numerical integration and after correction for floating potential size as indicated by EFI. Any digital quantity served will be provided an accompanying percentage error estimate. In certain regions of geospace it is known that RA moment quantities can not be accurate and fill data will be served for such intervals. As an example, when the mean energy of ions approaches the 10 keV, well below the upper energy limit of Hydra, no estimates of ion bulk flow will be computed, since there is substantial unmeasured momentum outside of Hydra's range. For that same spectrum we may be able to estimate the density and pressure, however. The RA will run in a batch mode, taking a request and processing it, and notifying the user by email when it is ready for access. The RA runs software that uses both the efficiency of FORTRAN and the Interactive Display Language IDL to "serve" requested plots or digital data according to the VWO protocols.

Robert Strangeway/University of California, Los Angeles

Value Added Services for VxOs: Creation of a Comprehensive Data Set for the FAST Small Explorer

The Fast Auroral Snapshot Small Explorer was turned off and no longer acquired data as of May 2009. Since its launch in 1996 the spacecraft has acquired nearly 13 years of data, with over 50,000 orbits. The spacecraft included a comprehensive suite of instruments

designed to measure the particles and fields within the Earth's auroral acceleration region. Having completed the data acquisition for the FAST mission it is appropriate to create a comprehensive data set that includes the best calibration data, as well as enhancing the data quality by, for example, removing the signatures of spacecraft nutation in the magnetic field data that are present after a spacecraft attitude maneuver. The data set will include definitive spacecraft attitude and ephemeris, definitive instrument calibration, a summary of the data acquisition and instrument modes for the entire mission, and the generation of the final data products for distribution. The data will be stored and distributed in a format used by the space physics community at large (e.g., CDF) via a web-server at University of California, Berkeley and a mirror site at the University of California, Los Angeles. The effort will collaborate closely with the UCLA node of the Virtual Magnetospheric Observatory, thereby ensuring compliance with the SPASE data model. The proposing team consists of the lead investigators for the experiments on FAST, who have the necessary knowledge and expertise to create a comprehensive and definitive set of the FAST data. These data will significantly enhance the science return, allowing many other investigators to carry out both detailed investigations of specific auroral phenomena, and large-scale statistical studies such as the distribution of magnetic fields perturbations and the ion and electron fluxes over the polar ionosphere for more than one solar cycle.

**Jon Vandegriff/Johns Hopkins University Applied Physics Lab
Value Added Services for VxOs: VO Downloader and Heliolib - a Tool and a
Library for Science Data Integration**

We propose to develop VO Downloader, a tool that will greatly simplify the downloading of data from Heliophysics Virtual Observatories. The focus of our efforts is to reduce the tedium of converting diverse data products into something that can be incorporated into a custom data analysis project. The VO Downloader will take a VxO query result (such as a list of URLs) and allow scientists to acquire arbitrarily large or small amounts of this data (less than one file or the entire dataset) in a format and layout of their choosing. Our tool will also allow content from two or more datasets to be merged in a science-sensitive way to create a custom, derived dataset. The tool will be based on an underlying data access library which we call Heliolib. This library will offer a single API for semantic access to any Heliophysics timeseries dataset, and it is aimed at ending the formats problem. The library will be the result of merging two existing mechanisms for uniform data access, Autoplot and DataShop, which have been developed separately but have both proven useful in providing access to distributed, remote data. In addition to being used within the VO Downloader, Heliolib itself will be made available in several languages (Java, IDL, Matlab) in order to allow scientists to write their own custom download routines. This effort is endorsed by seven VxOs: VHO, VEPO, VMO/G, VMO/U, VWO, ViRBO, and VITMO. Also, the downloader will be available from VSPO and will be incorporated into CDAWeb. The longer term maintenance of the VO Downloader will be a team effort managed by ViRBO, VHO, and JHU/APL, all of which are expected to be involved in VxO efforts for the foreseeable future. Virtual observatories are providing excellent ways to search for and locate data, and the VO Downloader will make it easier

to transform the search results from a VxO into a data product that is immediately usable and easily integrated with other datasets or with a scientist's own data analysis workflow.
