

The Past, Present, and Future of the NASA Heliophysics Data Environment (HPDE)

D. Aaron Roberts

NASA GSFC

HDMC* Project Scientist

Program Scientist, HPDE (HQ role from GSFC)

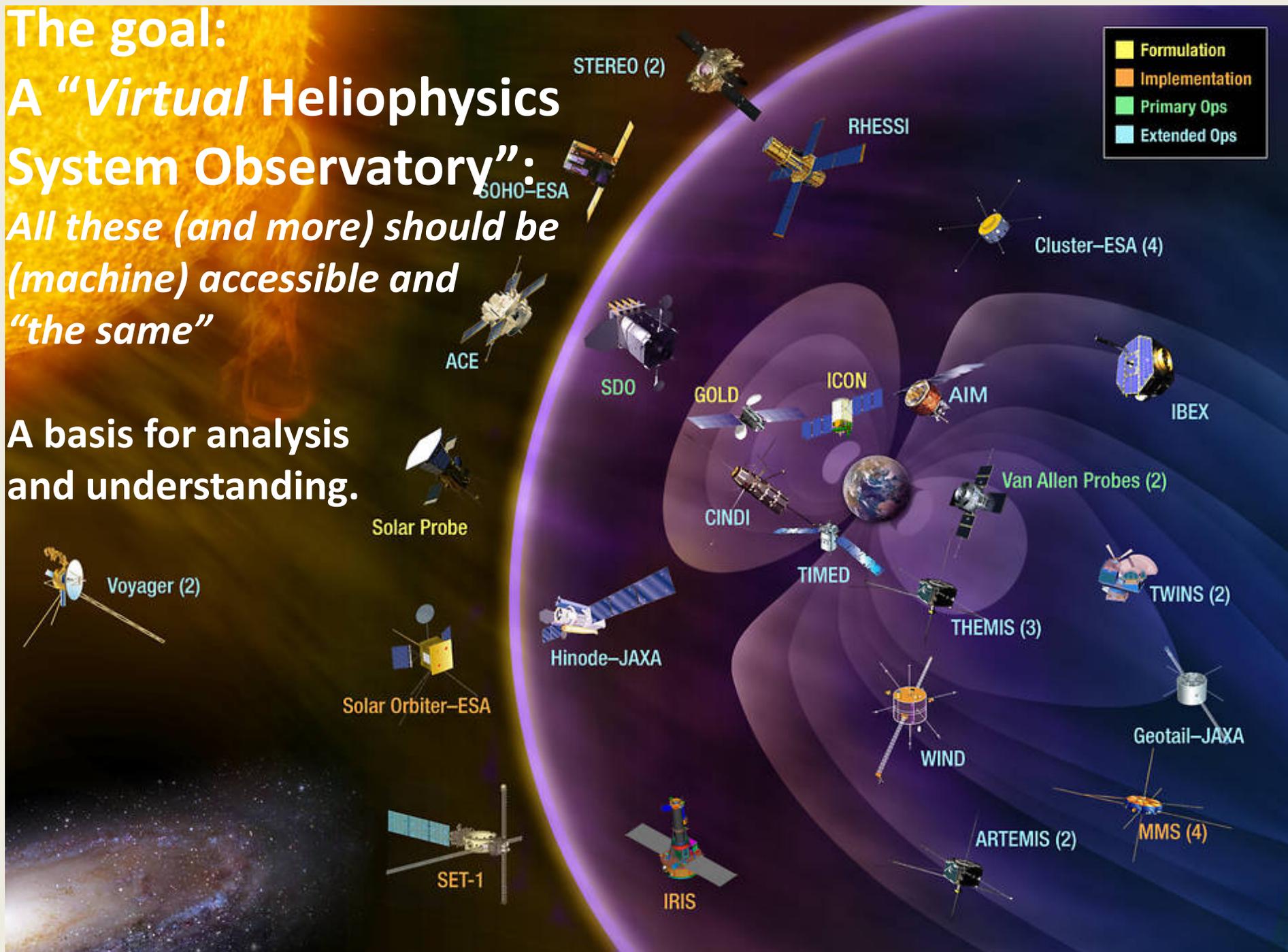
*Heliophysics Data and Model Consortium

The goal:

A “Virtual Heliophysics System Observatory”:

All these (and more) should be (machine) accessible and “the same”

A basis for analysis and understanding.



Overview

- *Brief history of the HP Data Environment*
- *Current Status*: Archives, Registry, Web (browser, ftp) and Web Service Access, Tools
- *Role of standards*
- *Current standards-based tools*
- *Remaining challenges*

See <http://hpde.gsfc.nasa.gov> (includes the Data Policy) and <http://heliophysicsdata.gsfc.nasa.gov> (generic data access via the “Heliophysics Data Portal”)

The “CODMAC”
report set the
course early on ...

Data Management and Computation
Volume I: Issues and Recommendations
Committee on Data Management and
Computation, Space Science Board,
Assembly of Mathematical and Physical
Sciences, National Research Council
ISBN: 0-309-12330-5, 185 pages, 8 1/2 x
11, (1982)

7. PRINCIPLES FOR SUCCESSFUL SCIENCE

DATA MANAGEMENT	135
I. Scientific Involvement	135
II. Scientific Oversight	137
III. Data Availability	137
IV. Facilities	138
V. Software	138
VI. Science Data Storage	139
VII. Data-System Funding	139

NASA

What scientists want to do



- Model physical processes
- Study physical interactions
- Use multi-source data



NASA

What scientists don't want to do



- Wonder if they have all the data available
- Wait a long time to get the data
- Spend time and money getting data into a useable format
- Spend time and money fighting computers



NASA

What scientists want a data system to do



- Locate relevant data
- Access data quickly
- Create versatile data sets
- Use the data easily



NASA

What data systems managers want scientists to do



- Hold to agreed upon standards
- Submit all relevant data
- Document all relevant data
- Report location of relevant data



Viewgraphs (!) from the first "Space Physics Data System" meeting in 1990. Our needs have been remarkably constant. We now do this!

History

- *Brief history of the HP Data Environment:*
 - From proprietary “lab benches in space” to an *open*, comprehensive, searchable, accessible, distributed archive.
 - From “give something to NSSDC” as an archiving plan to a coherent **NASA HP Science Data Management Policy** (orig. 2007; v1.1 2009, Archives codified; v 1.2 2016, consolidate “VxOs”).
 - Current distributed archives: discover and access data from **hundreds of observatories** (spacecraft, etc.) covering **more than 50 years** of exploration; many (most of NASA) datasets are at the HP (active) archives SPDF or SDAC at GSFC.
 - **Data/model integration** is increasingly important.
 - Current services and development support the “**DRIVE**” initiative of the NRC **Decadal Survey**.

NASA Heliophysics Science Data Management Policy

Orig. release 6/25/07; latest version 10/4/16

“The success of this effort depends on having **scientific involvement in all stages** [3, 4] of data production, dissemination, and archiving, with a close collaboration between scientific and technical teams. Two overarching principles also essential to achieving the goals of current Heliophysics programs are:

1. Embracing NASA’s **open data policy** that high-quality, high-resolution data, as defined by the mission goals, will be made publicly available as soon as practical, and

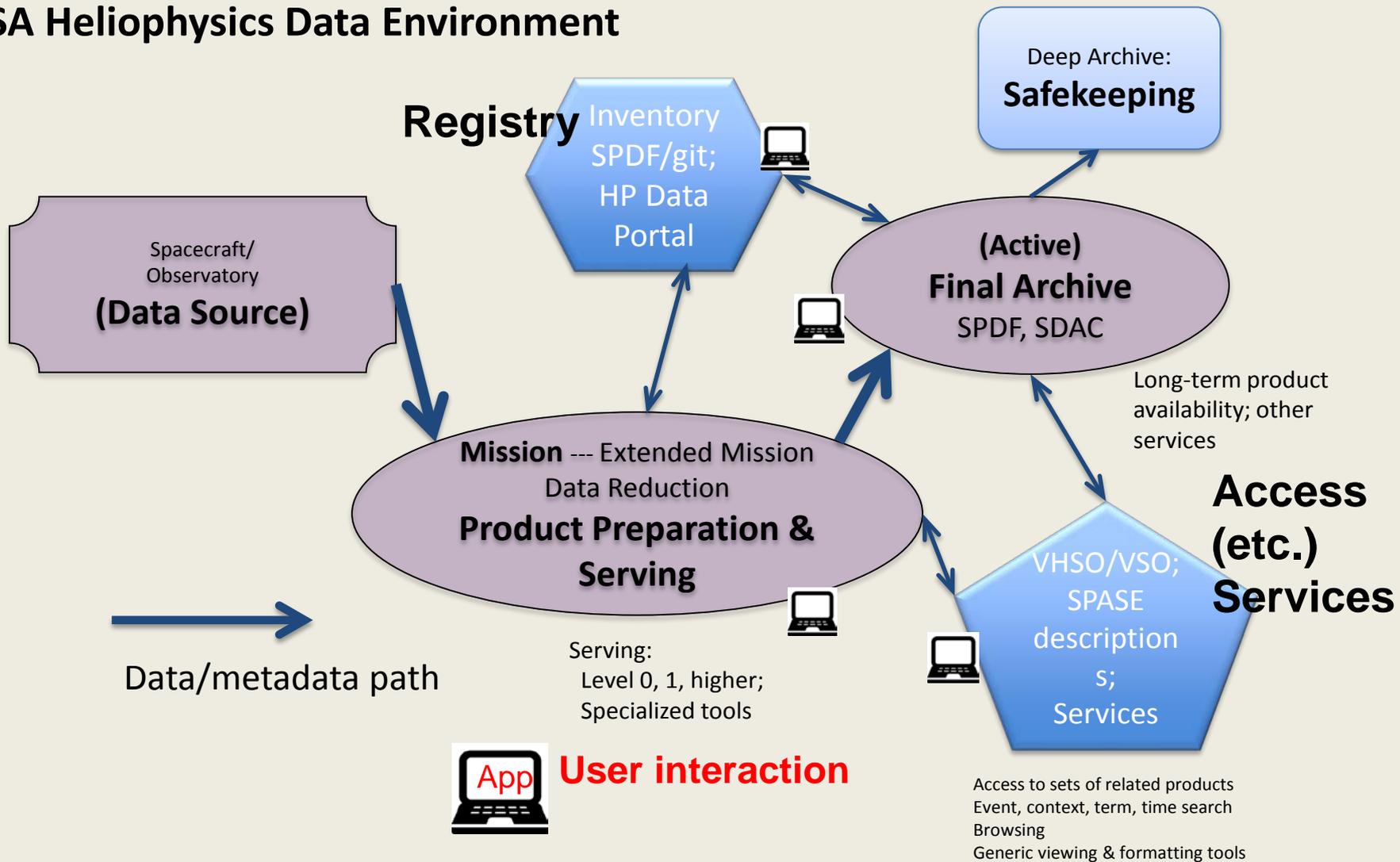
2. Adhering to the goal of **early and continuing independent scientific data usability**, which requires uniform descriptions of data products, adequate **documentation**, sustainable and **open data formats**, **easy electronic access**, appropriate **analysis tools**, and care in data **preservation**.

This view involves the data users from the general science community as responsible partners in the improvement of the data environment and of the data products themselves”

Decadal Survey: Overview

- Significant progress has been made over the last decade in establishing the essential components of the solar and space physics data environment. However, to achieve key national research and applications goals, *a data environment that draws together new and archived satellite and ground-based solar and space physics data sets and computational results from the research and operations communities is needed.*

NASA Heliophysics Data Environment



SPDF = Space Physics Data Facility **SDAC** = Solar Analysis Data Center
Both at GSFC; official HP Final Archives, both actively provide data

Virtual Observatory

- A “virtual observatory” is a system that provides a general capability to **perform queries** for datasets and to **access** them when found.
 - Ideally, we want all the variegated datasets from diverse observatories to look “the same” making analysis enormously easier. (*Virtual Heliophysics System Observatory*)
- The use of **standards** allows **applications** and **interoperable tools** to share data transparently. A VO is **the invisible machinery that allows us to work efficiently**.
- Format, metadata, and data-access API standards are all needed.

STANDARDS, STANDARDS, STANDARDS

- **Data format** (FITS, CDF, NetCDF): Self-describing, with standard metadata for easy reading by **standard tools**.
- Basic registration **metadata** terms and format (SPASE): Easy access to all datasets; inventory tracking.
- **General API** for data **access** (HAPI): Let all tools have easy access to data, independent of underlying format.

Standards-Based Tools

- Standards have allowed us to implement a number of “value-added services”:
 - A general product finder and accessor (Heliophysics Data Portal)
 - CDAWeb (Serving Space Physics archive and active mission data)
 - General analysis tools (“SPEDAS” and “SolarSoft” with graphical and other tools and direct data access via VSO, CDAWeb, and other APIs)
 - A viewer for a wide variety of solar images with overlays, movies, and events (Heliviewer)
 - Data/model comparison tools (at CCMC; “VMR” at U Michigan)
 - Other search engines, visualization tools, etc.
- We need further integration of datasets, models, analysis tools, and visualization.
- *The examples that follow are largely made possible standard metadata, data formats, and web services.*



Heliophysics Data Portal

“Find it. Browse it. Get it.”

[Help](#)[Geo Orbits](#)[Helio Orbits](#)[SPASE Registry](#)[ADS Abstracts](#)[Weather](#)[Feedback](#)

Text Restriction

Time Span Restriction i

YYYY-MM-dd or YYYY-DDD

from:

to:

Element Restriction i

[Resource type](#) i

[Measurement type](#) i

[Observatory Group](#) i

[Observatory](#) i (~200!)

[Instrument](#) i

[Observed region](#) i

[Spectral range](#) i

[Cadence](#) i

[Repository Name](#) i

[Access rights](#) i

[Format](#) i

Current Product Restrictions

No restrictions are currently set.

Showing 1 - 20 of 1601 Results

[View Current List](#)

Sort by

Observatory ▾

Products (& SPASE descriptions)

Access Links

1	ACE 27-day Survey Plots	<ul style="list-style-type: none"> • Polar-Wind-Geotail 'gif-walk' site <input type="button" value="Get Images/Plots"/>
2	ACE Cosmic Ray Isotope Spectrometer (CRIS) 1-Hour Level 2 Data	<ul style="list-style-type: none"> • FTP access to files at SPDF • HTTP access to files at SPDF • CDAWeb • ACE Science Center FTP area • ACE Science Center • ACE Cosmic Ray Isotope Spectrometer (CRIS) Level 2 data Home Page <input type="button" value="Get Data/Plots"/>
3	ACE CRIS L2 1-day Z=3-28 flux data	<ul style="list-style-type: none"> • ACE Science Center • ACE/CRIS L2 data in HDF via ftp • CDAWeb • FTP access to files at SPDF • HTTP access to files at SPDF <input type="button" value="Get Data/Plots"/>
4	ACE Daily Survey Plots	<ul style="list-style-type: none"> • Polar-Wind-Geotail 'gif-walk' site <input type="button" value="Get Images/Plots"/>
5	ACE Electron Proton Alpha Monitor (EPAM) 1-hour Key Parameter data	<ul style="list-style-type: none"> • FTP access to files at SPDF • HTTP access to files at SPDF • CDAWeb <input type="button" value="Get Data/Plots"/>

Directly Read Data from CDAWeb into IDL

```
timename='jul_day' ;name of time variable -- Julian days  
start_time = '1998-06-10T00:00:00.0Z' ;start time  
stop_time = '1998-06-10T23:59:59.0Z' ;stop time  
dt_sec=10.0 ;sec -- bin size in seconds
```

```
dataset_id='WI_H0_MFI' ; CDAWeb dataset ID  
vars=[ 'B3F1=Bmag3', 'B3GSE=Bx3,By3,Bz3' ] ; CDAWeb variable names with locally assigned names  
cdaweb_get_bin, dataset_id,vars,start_time,  
stop_time,dt_sec,time_name=timename
```

```
dataset_id='WI_PM_3DP'  
vars=[ 'P_DENS=np3', 'P_VELS=Vxp3,Vyp3,Vzp3', 'P_TEMP=Tp3', 'A_DENS=na3', '$  
'A_VELS=Vxa3,Vya3,Vza3', 'A_TEMP=Ta3']
```

```
cdaweb_get_bin, dataset_id,vars,start_time,  
stop_time,dt_sec,time_name=timename,/autobad
```

No more writing code for every dataset.

The Internet functions as a local, easy to use hard drive.

Prototype for more general “Virtual Observatory” access to any HP data.

VSO does the same for Solar Data

“AMDA” (Europe) uses SPDF web services

cdaweb.gsfc.nasa.gov/istp_public/

STEREO-A Qu...ook Movies Raptor System CDAWEB_GET_BIN FramesByMail Apple Yahoo! Google Wikipedia News Popular LLWebmail

Python Object Oriented Living With a Star and Heliophysics Data Environment SPDF - Coordinated Data Analysis Web (CDAWeb)

**GODDARD SPACE FLIGHT CENTER**
Space Physics Data Facility

[+ Goddard Home](#)
[+ Visit NASA.gov](#)

[+ SPDF HOME](#) [+ DATA & ORBITS](#) [+ MODELS at CCMC](#) [+ SCIENCE ENABLED](#) [+ AND MORE](#)

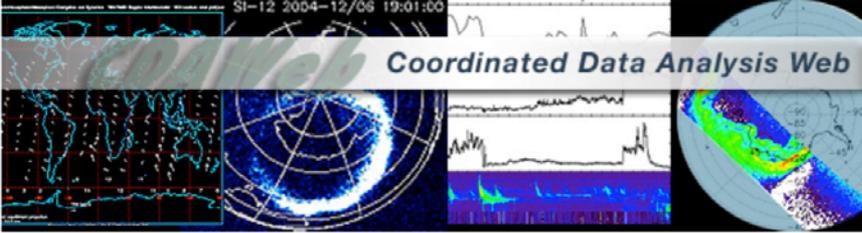
CDAWeb
[+ CDAWEB HOME](#)
[+ FEEDBACK](#)
[+ ABOUT CDAWEB](#)

CDAWeb Mirror Site
[+ RAL/UK](#)

Guides and Tutorials
[+ CDAWeb help](#)
[+ Internet browser help](#)

Additional Services
[+ CDAWeb Inside IDL](#)
[+ HTTP & Anonymous FTP access to public CDAWeb database](#)
[+ Overview of Alternative Data Access Methods](#)
[+ Autoplot.org \(non-NASA\) interface to public CDAWeb database](#)

Additional Resources
[+ Usage Statistics](#)
[+ GIFWALK Data and Orbit plots](#)
[+ Space Physics Use of CDF](#)
[+ Data Inventory Graph](#)
[+ Home Pages for ISTP Investigations](#)



Coordinated Data Analysis Web

Coordinated Data Analysis Web (CDAWeb)

Public data from current space physics missions

NEW
Summer 2016: New Level 2 MMS data sets and data coverage continue to be added to the system.

PRIOR DATA & SOFTWARE UPDATES ...

- **Select one OR more Sources**
(default = All unless no Instrument Types selected)
- **AND Select one OR more Instrument Types**
(default = All unless no Sources selected)

ACE
 ARTEMIS
 BARREL
 CNOFS
 CRRES
 Cluster
 Cubesats
 DMSP
 Equator-S
 FAST
 GOES
 GPS

Activity Indices
 Electric Fields (space)
 Electron Precipitation Bremsstrahlung
 Engineering
 Ephemeris
 Ephemeris/Attitude/Ancillary
 Gamma and X-Rays
 Housekeeping
 Imaging and Remote Sensing (ITM/Earth)
 Imaging and Remote Sensing (Magnetosphere/Earth)

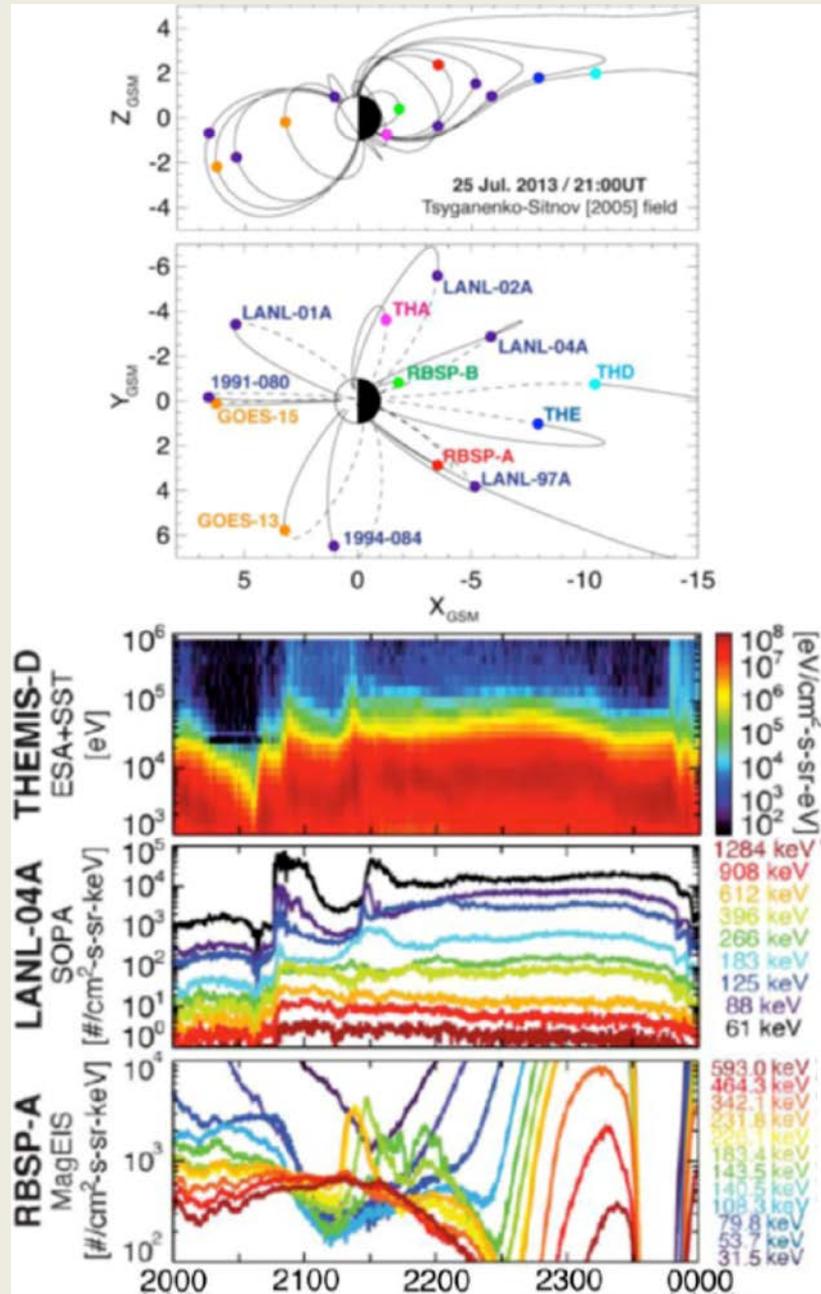
SPEDAS (Space Physics Environment Data Analysis Software)

spedas.org

Generic IDL analysis tool with a stand-alone version and with direct links to much of Space Physics data, including all of SPDF.

Roughly the Space Physics equivalent of SolarSoft (supported via SDAC).

Developed by various missions, most recently THEMIS. Used now by MMS and many others.



Helioviewer: Multispacecraft solar imaging and movie-making

The screenshot displays the Helioviewer.org website interface. At the top, the browser address bar shows "helioviewer.org". Below the browser, there are navigation links for various services like Raptor System, CDAWEB_GET_BIN, FramesByMail, Apple, Yahoo!, Google, Wikipedia, YouTube, News, Popular, LLWebmail, DC_Baltimore_Traffic, Apple, Amazon, HDP, and HPDE. A search bar contains the text "Monitor Your Credit Score With Credit Tracker | Capital One".

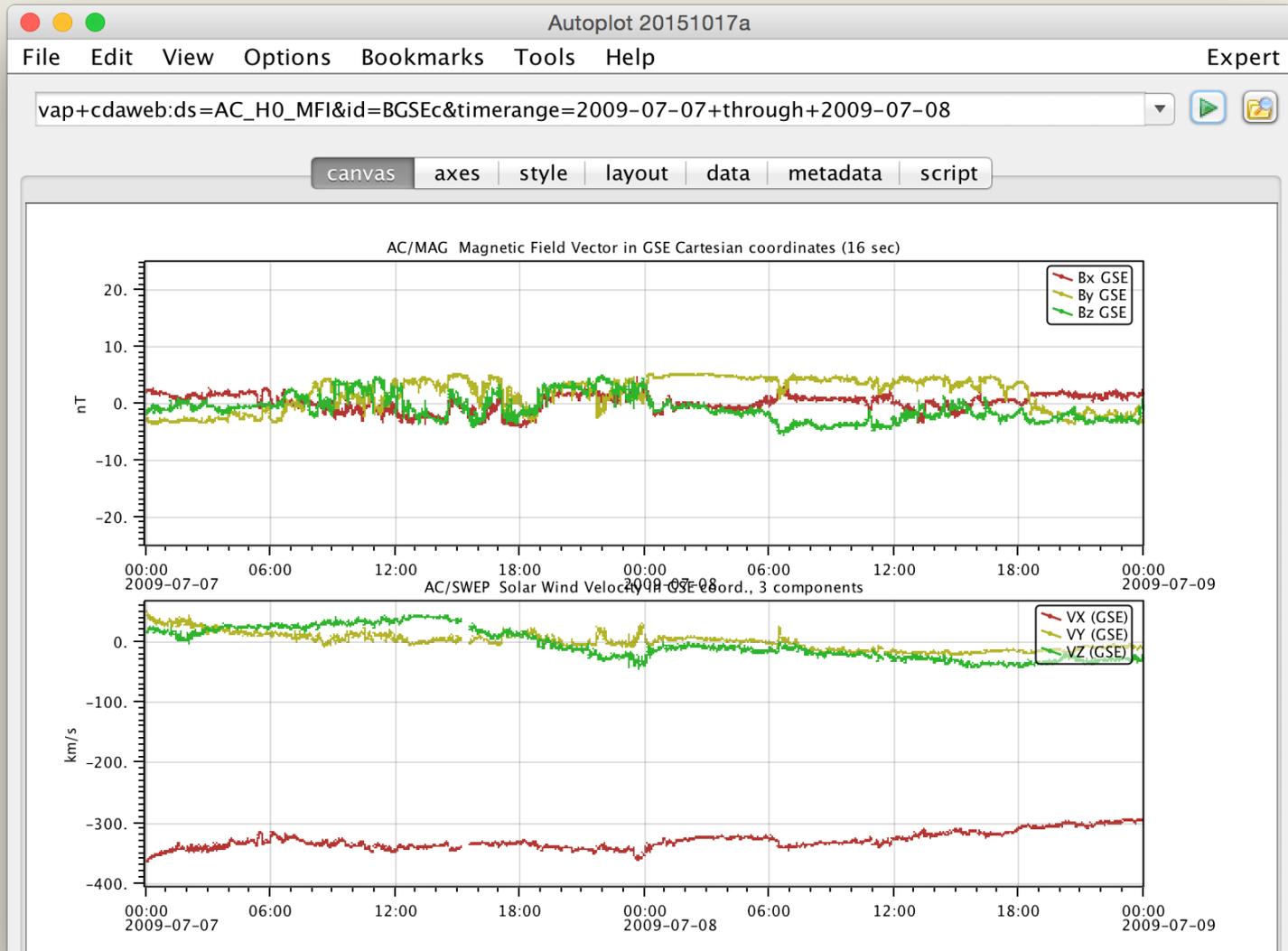
The main content area is divided into several sections:

- Time Selection:** A panel on the left allows users to select a date (2015/10/27), time (14:32:02 UTC), and time-step (1 Hour).
- Images:** A list of recent images is shown, including LASCO C2, EIT 195, and LASCO C3.
- Solar Features & Events:** A section for filtering solar features, with a list of categories such as Active Regions, Coronal Cavities, Coronal Dimmings, Coronal Holes, Coronal Jets, CMEs, Coronal Rains, Coronal Waves, Emerging Fluxes, Eruptions, Filaments, Filament Activations, Filament Eruptions, Flares, Loops, Oscillations, Plages, Sigmoids, Spray Surges, and Sunspots.
- Central Image:** A large, circular solar image showing the Sun's corona in red and orange, with a green Earth-like disk in the center. The image is labeled "center" and has a vertical slider for zooming.
- Right Sidebar:** Contains "Helioviewer Project News" with several announcements, a "YouTube" logo, and "User-Generated Movies" with a video thumbnail.

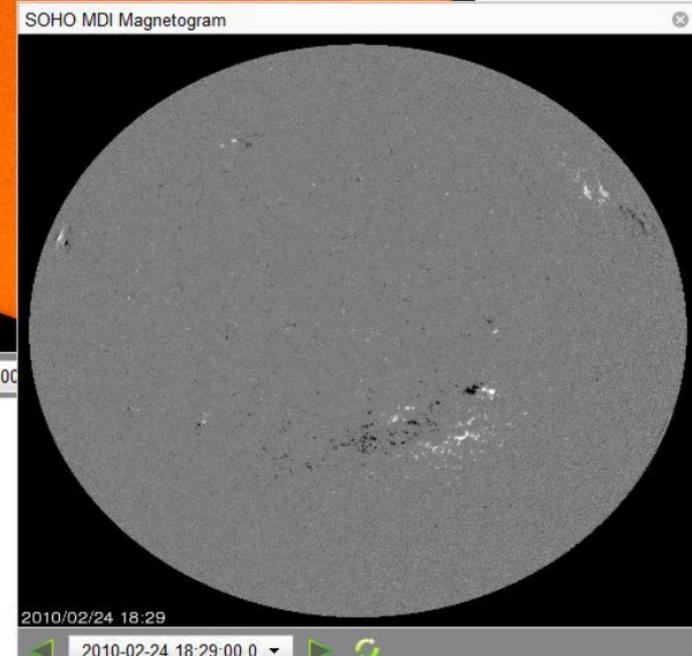
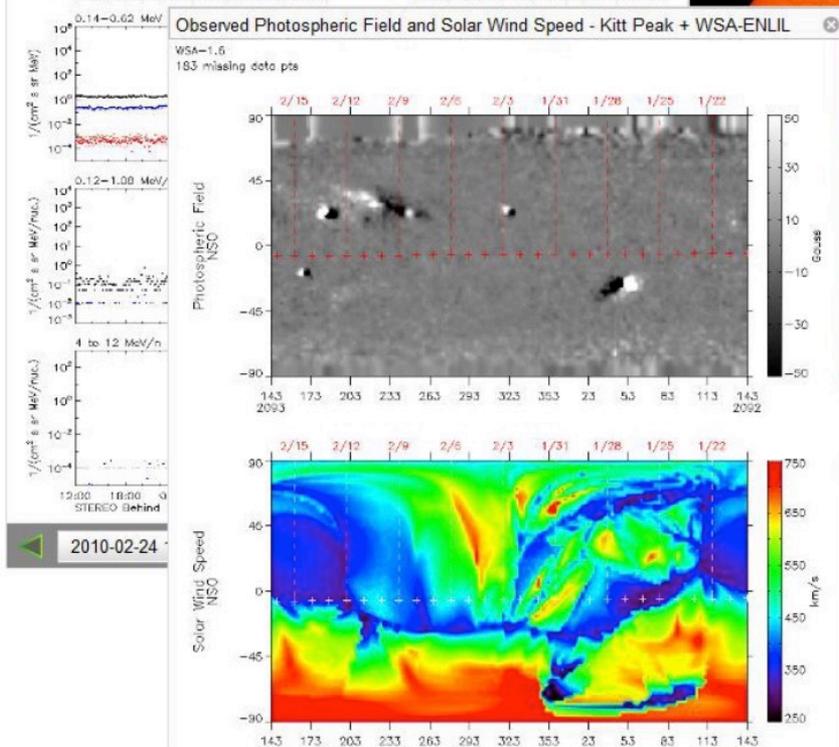
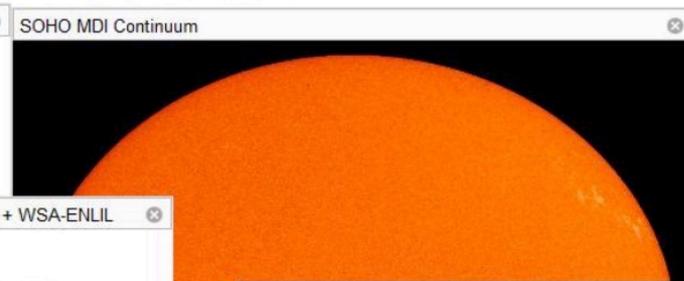
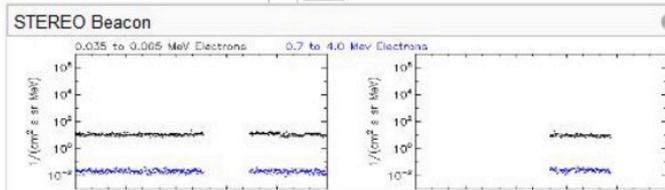
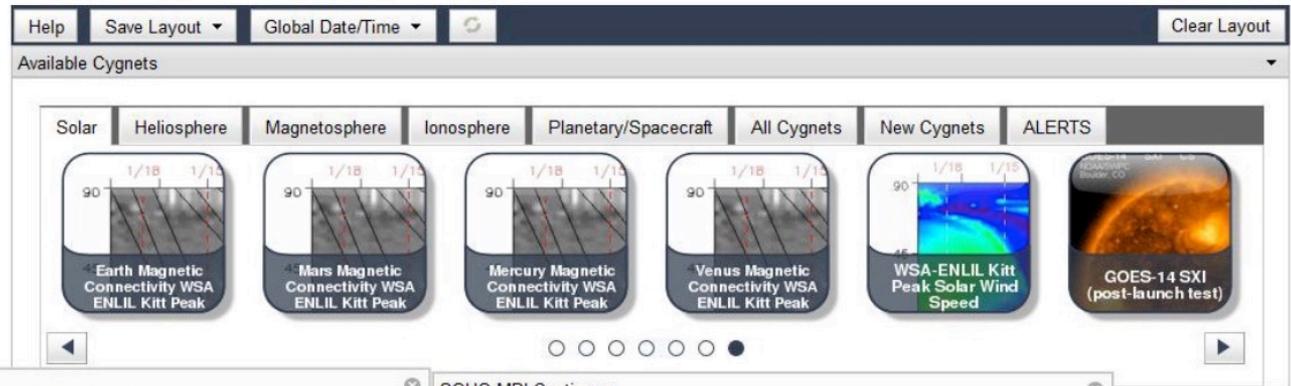
At the bottom of the page, there is a footer with navigation links: Help, Glossary, About, Usage Tips, Wiki, Blog, JHelioviewer, Public API, @helioviewer, Contact, and Report Problem.

Autoplot (autoplot.org): Reads multiple formats including [Complex ASCII tables](#); [Binary tables](#); [Common Data Format \(CDF\)](#); [NcML](#); [SPASE](#); [Cluster Exchange Format](#); [NetCDF](#); [OpenDAP](#); [HDF5](#); [TSDS](#); [FITS](#); [Excel](#); [Wav](#); [PNG](#), [JPG](#), etc. For details and a full list, see [Formats](#).

Makes “gif-walk” summaries; mutli-dataset overviews, etc.



Integrated Space Weather Analysis (iSWA) at CCMC: Observed and simulated near real-time data

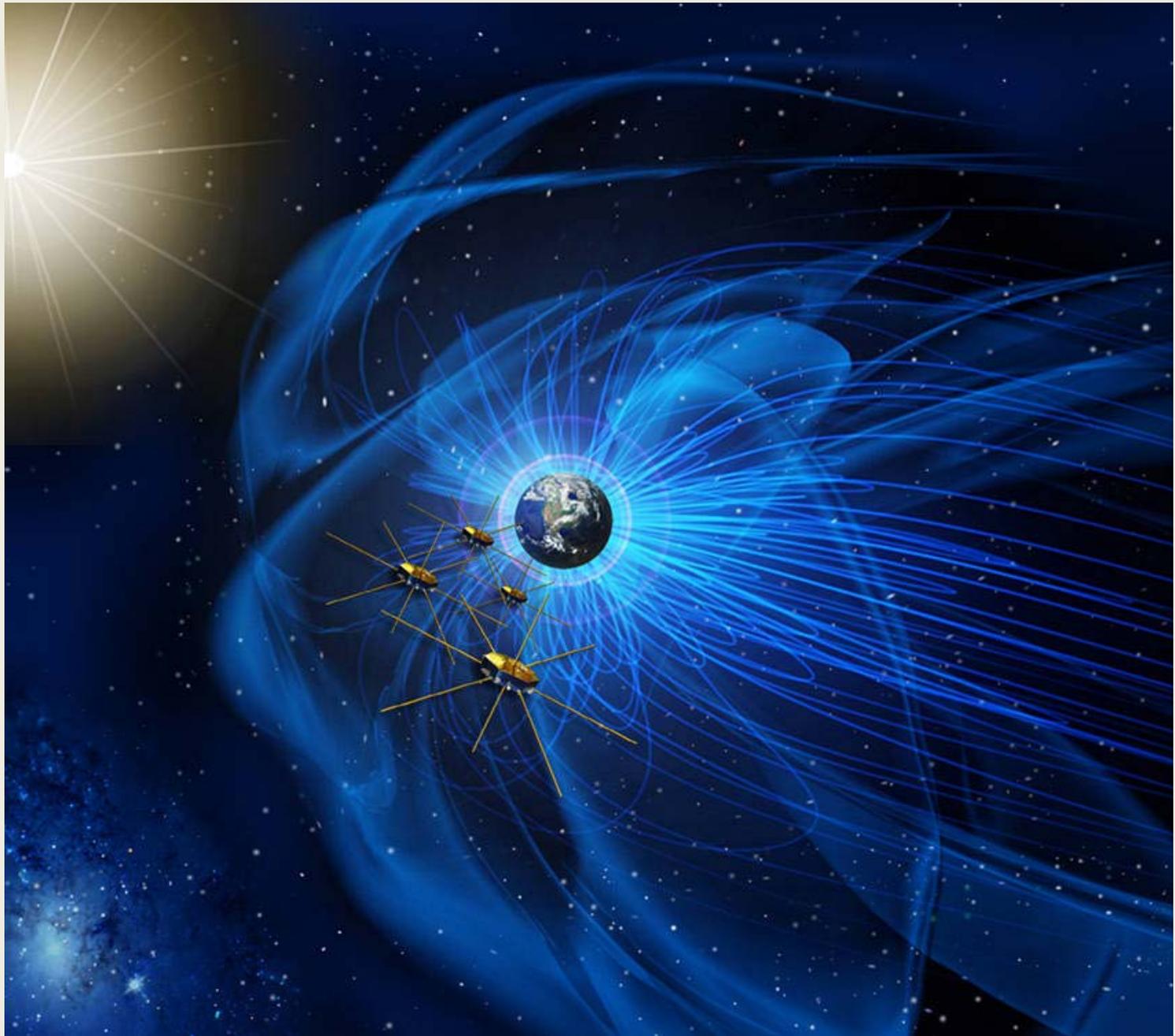


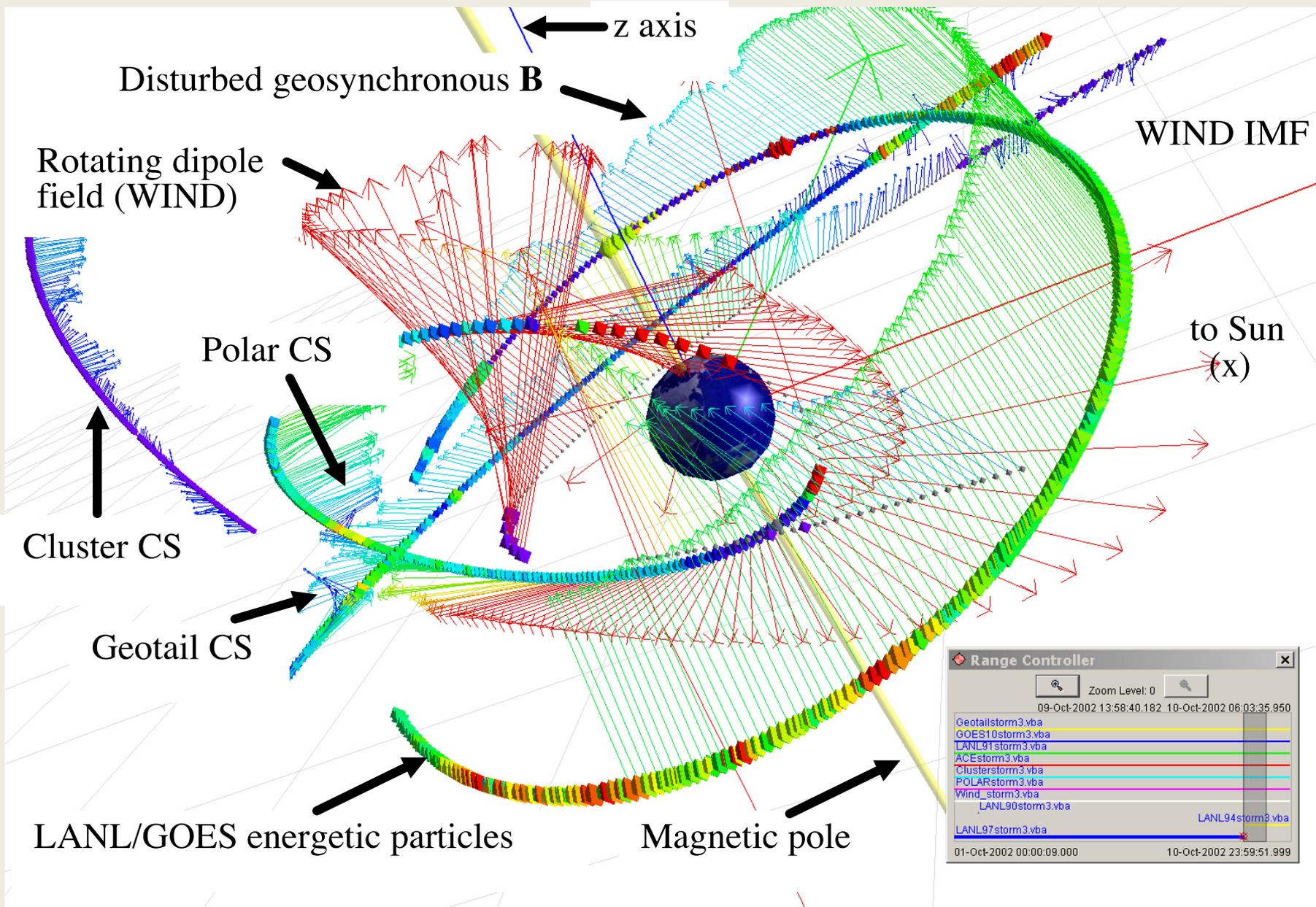
Decadal Survey: Future Goals

- Heliophysics is poised to make a natural transition from being driven predominantly by the pursuit of basic scientific understanding of physical processes towards one that must also address more operational, application-specific needs, much like terrestrial weather forecasting. *This transition requires **(1) instant unfettered access to a wide array of datasets from distributed sources in a uniform, standardized format, (2) incorporation of the results of community-developed models, and (3) the ability to perform simulations interactively and to couple different models to track ongoing space-weather events.*** [(3) is more CCMC]
- NASA has already taken the important first step in integrating many of these datasets and tools to form the Heliophysics Data Environment (HPDE). *The main objective of the HPDE is to implement a distributed, integrated, flexible data environment. HPDE modeling centers should serve as a sound foundation for a future, **fully integrated heliophysics data and modeling center.***

Some Major Future Challenges

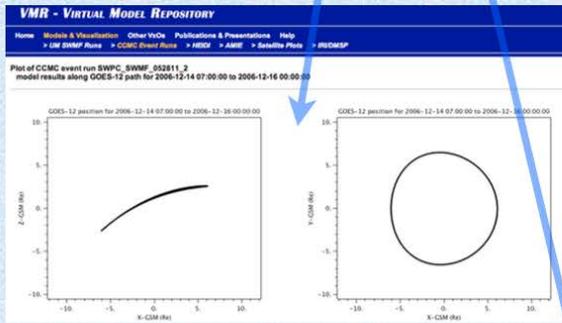
- **Large data volumes and new data analysis methods**
 - How to use the data: *Pattern recognition; data mining; 4-D visualization*; use of *modern data analytics methods*: clustering, neural nets, “deep learning” etc.
 - No magic wands, but great efficiencies possible
 - Requires work to merge disciplines
 - Physical insight is difficult; *feature/process identification and prediction are lower-hanging fruit.*
- **Model-data comparisons and insights**
 - *Seamless integration of model output with data streams*
 - Data assimilation; *true space weather capabilities*





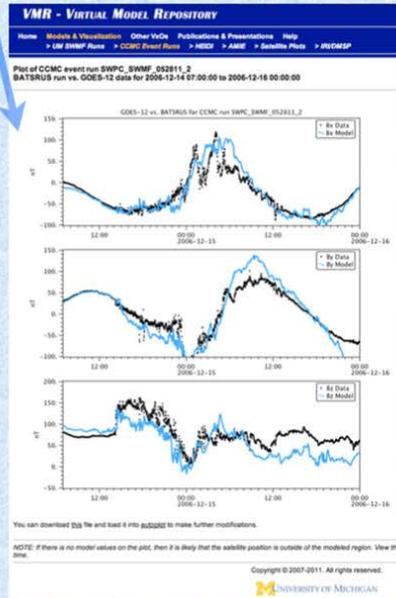
Multiple plot types are available:

- satellite B field
- model output along satellite trajectory
- data/model comparison
- satellite location



The data/model plot on the right includes many parts:

- model run (CCMC)
- satellite trajectory (SSCWeb)
- model interpolation (CCMC)
- satellite data (CDAWeb)
- collect plot pieces (VMR)
- plotting (Autoplot)



<http://vmr.engine.umich.edu>

Now also at CCMC

Data-Model Comparisons

Current HDMC effort: “Virtual Model Repository”

Uses Internet connectivity to load spacecraft data and simulations into the same visualization.

A way forward:

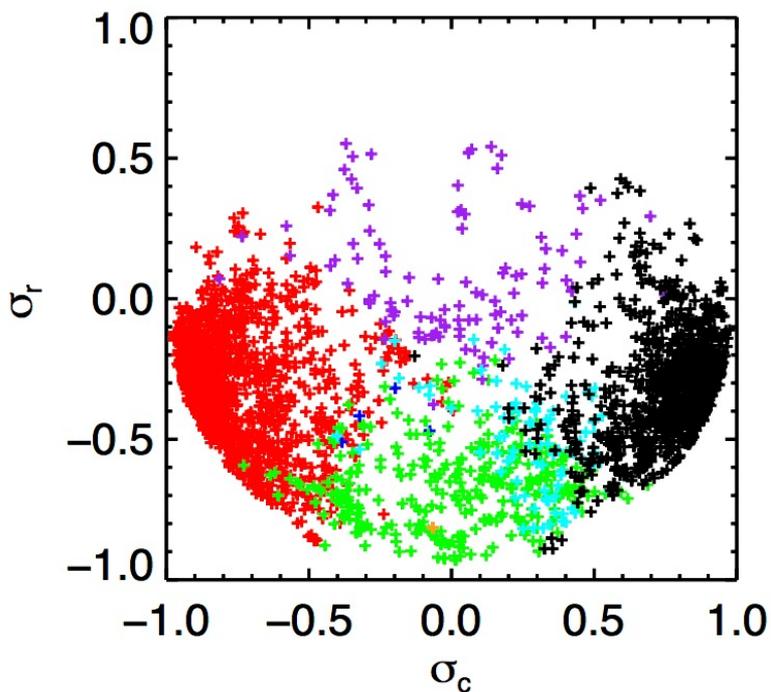
Develop a community-based plan for generalizing this work.

Start with the CCMC to develop ideas and come up with a cohesive approach.

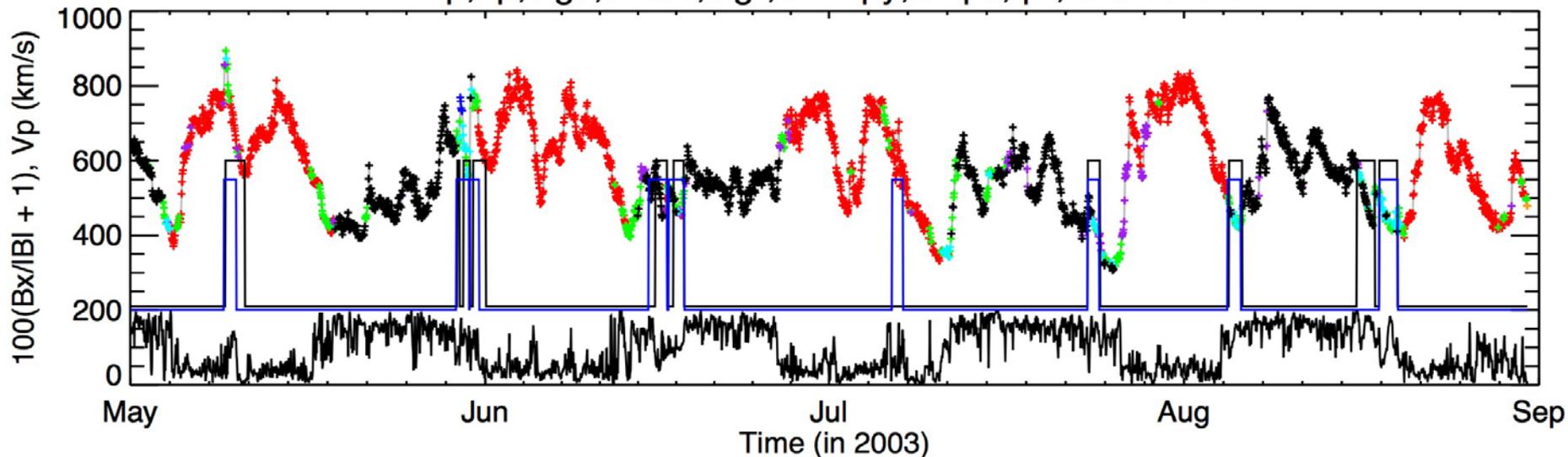
Fund specific efforts to implement the plan.

Solar Wind States

Finding **clusters** of solar wind states, given by the values of speed, temperature, density, and derived quantities such as “cross helicity” (**8-D space**). A rapidly convergent algorithm gives “natural” or “objective” states of the solar wind. [“k-means clustering”]



vp,np,sigc,o7to6,sigr,entropy,avqfe,tpr; 8 Clusters



Questions from the Committee

(a) What is the future-looking direction of the HPDE (and where it's starting from)? [See above]

(b) What programs, resources, or partnerships are currently in place (or would be helpful) to move the program forward?

- CCMC, NSF, ESA, IMPeX, Japan, Canada, ISWI, DataCite, PDS/PPI, EOSDIS; distributed observatories and tools
- Need (and expect) continued support from NASA HQ; “DRIVE” enhancements should help

(c) How might the community better utilize or take advantage of this resource?

- Marketing, training, education, cool examples, ...
- Better partnerships, more formal NASA/NSF/ESA ... links
- Data Policy tweaks
- Continued community resource development

