# Evolution of Data Policies and Practices within NASA's Heliophysics Program during Solar Cycle 23

Charles P Holmes
NASA HQ "*emeritus*"

December 15, 2008

This study was supported by the eGY

# Accomplices

Tom Armstrong - FTC
Scott Bailey - VT
Joe Bredekamp - NASA HQ
Geoff Crowley - ASTRA
Joe Giacalone - U. of Arizona
Leon Golub - CFA
Joe Gurman - GSFC
Todd Hoeksema - Stanford
Russ Howard - NRL
Hugh Hudson - UCB
Janet Kozyra - U of Michigan
Lou Lanzerotti - NJIT

Janet Luhman - UCB
Tom Moore - GSFC
Bill Peterson - eGY, U of Colorado
Joleen Pickett - U of Iowa
Aaron Roberts - GSFC
Karel Schryver - LMSAL
Jim Sharber - SWRI
Ruth Skoug - LANL
Chris St Cyr - GSFC
Ray Walker - UCLA
Thomas Zurbuchen - U of Mich

# Motivation

- This paper will trace the evolution of the Heliophysics data practices and present the results of a survey to ascertain the benefits and problems of undertaking the transition.
- My goals are to point out the contributing factors to the evolution of the data environment and make an assessment about the added value that the evolution made to Heliophysics science.
- The causes for this shift in paradigm were many:
  - the IT revolution [the rise of the ubiquitous internet and Moore's law effects on hardware and software],
  - Shifts in attitudes about the roles of the principle investigators (PIs) in sharing data from their investigations,
  - Directions from NASA HQ, and
  - Other.

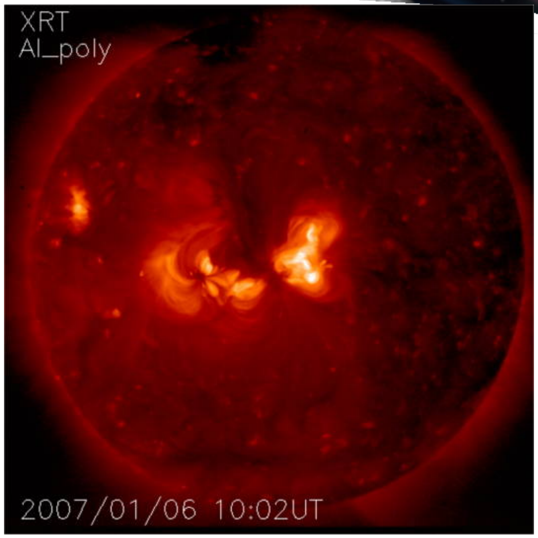
# The Heliophysics Science Disciplines

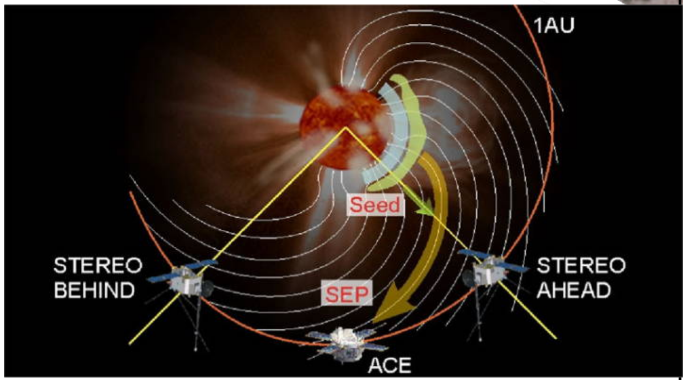
Solar

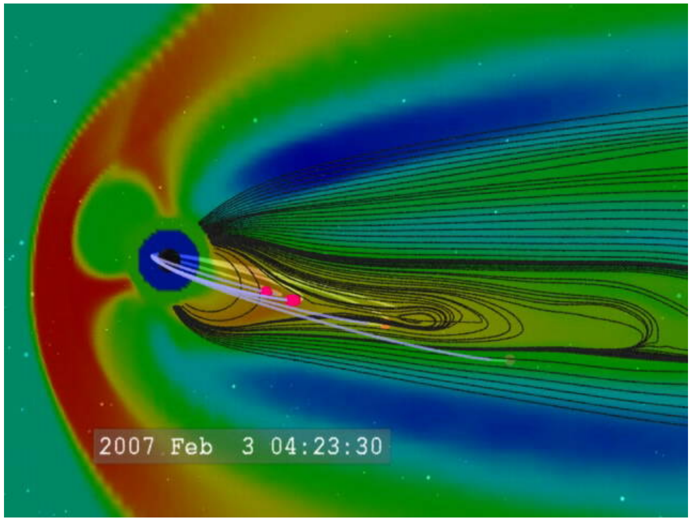
Heliospheric

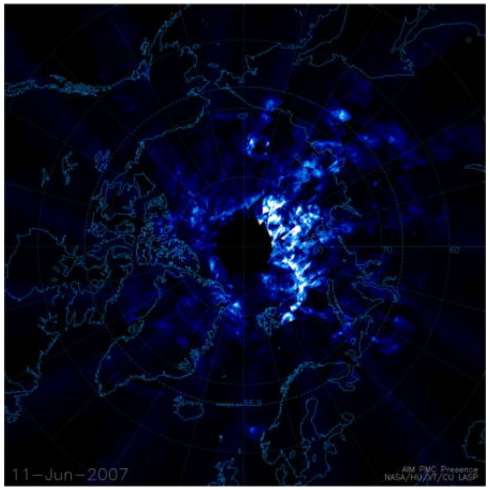
Magnetospheric

ITM



# Short History

- To set the stage, I am comparing the data environment of the mid-1990s to that of today.
  - This covers roughly Solar Cycle 23 and, coincidently, my tenure at NASA HQ where I was privileged to work closely with the ongoing heliophysics investigations enabled by the operating fleet of satellites and instruments.
- In the mid-90s the ISTP fleet had just been launched along with SOHO to complement Ulysses, Voyager, Yohkoh and the early SMEXes: SAMPEX and FAST.
- The start point of my comparison is prior to the launches of ACE, TRACE, IMAGE, TIMED and RHESSI.

# The Science Missions in 1996 and 2008
# Spanning Solar Cycle 23

| 1996 | Heliophysics Observatory | Great (2008) |
|---|---|---|
| SOHO (S & H) | SOHO (S & H) | Cluster (M & H) |
| Yohkoh (S) | RHESSI (S) | Geotail (M & H) |
| IMP 8 (H) | Hinode (S) | FAST (M) |
| Pioneer 10 (H) | STEREO (S & H) | Polar (M) |
| SAMPEX (H & M) | TRACE (S) | THEMIS (M) |
| Ulysses (H) | ACE (H) | TWINS (M) |
| Voyager (H) | IBEX (H) | AIM (ITM) |
| Wind (H) | Ulysses (H) | C/NOFS (ITM) |
| Geotail (M&H) | Voyager (H) | TIMED (ITM) |
| FAST (M) | Wind (H) | |
| Polar (M) | | |

Also, the magnetospheric IMAGE mission operated from 3/2000 to 12/2005.

# The Heliophysics Data Environment

- Data from the Heliophysics (HP) Great Observatory reside in a distributed environment and are served from multiple sources:
  - Multimission data centers located at Goddard SFC.
  - Mission-level active archives: e.g. ISTP, SOHO, ACE, TIMED, TRACE, IMAGE, Cluster, STEREO, THEMIS, etc.
  - Much of HP data are served from individual instrument sites.
- HP is evolving their new data environment with
  - Virtual Observatories for convenient search and access of the distributed data, and
  - Resident Archives to retain the distributed data sources even after mission termination.
- There is a Data and Computing Working Group to help the policies and implementations move ahead.

# The HP Data Policy

- In 2007, HP approved its Science Data Management Policy to improve management and access of HP mission data.
- Basic Philosophy - Evolve the existing HP data environment:
  - take advantage of new computer and Internet technologies to respond to the evolving mission set and community research needs (enable the HP Great Observatory).
  - Assure that the HP science community participate in all levels of data management.
- Guiding Principles
  - All data produced by the HP missions will be open and made available as soon as is practical.
  - Data will be independently scientifically usable [documentation, easy access, analysis tools].
- Distributed architecture employing the VxOs and SPASE as integrating forces.
- Implementation employing peer reviews and user community feedback.
- See http://hpde.gsfc.nasa.gov/ for more information.

# Survey of HP Scientists

- I made survey of many of the HP scientists who forged new ways of distributing, sharing and analyzing heliophysics data over the past solar cycle.
- I was particularly interested in the views as to the relative importance of the factors and welcomed candid comments and perspective.
- The survey questions were (summaries of responses follow):

A. Have there been significant changes in how you access, distribute and work with Heliophysics data today as compared to the mid-1990s?

B. Did you see changes in the ways that collaborations may have worked and/or the approaches to incorporating multiplatform data?

C. What do you think brought about some of these changes that you have experienced?

D. What was your involvement in the evolution of the Heliophysics data environment during this period?

E. Have the changes in the data environment promoted advances in Heliophysics science?

F. Other comments on this topic?

# A. Significant Changes since mid-1990s?

- Yes, yes, yes, …!
- mid-1990s:
  - PI "ownership" of data with limited distribution outside of instrument team.
  - Single-point, single-instrument investigations.
  - Data distribution by mail: data tape or CDs or by ftp from "flat" directories.
- Now:
  - PIs as stewards of the data sets.
  - Open data sets with rapid availability.
  - Web-based [browser, API, web-objects] access.
  - Feedback to PI: hundreds of eyes see more than two.
  - Multi-platform analyses of extended, complex phenomena.

# B. Changes in Workings of Collaborations and Multiplatform Studies?

- Yes.
- Multi-platform data are easier to acquire.
- Methods for examining, analyzing and displaying large quantities have become impressively functional and useful in the past five years.
- More rapid discoveries across a wider research community - put not as deep and thorough.
- It is much easier for remote colleagues to be sure that they are looking at identical views of data sets.

# C. What Brought These Changes?

- Technology - web, faster more capable processors, storage revolution.
  - The expansion of the internet and with that readily available commercial products and tools for the aspects in the data chain: observation -> analysis and discovery.
- Program guidance from HQ.
  - NASA and a new generation of PIs made the right decisions at a time when technology was making the identification of access to open data much easier.
- Cultural shift: from PI-ownership of the data to community ownership with the PI as the steward.
  - Adopting open data lead to change in the community behavior and provided positive feedback to the data producer -- more eyes and brains inspecting the data.
- The role of data standards.

# D. What Were Respondents' Involvement in the Data Evolution?

- Respondents spanned the HP disciplines.
- Representatives are
  – From instrument teams,
  – Data Analysts,
  – Modelers, and/or
  – Members of science advisory and proposal review panels.
- All witnessed, experienced, advocated, and benefited from the evolution.

# E. Has the Evolution Promoted Advances in Heliophysics Science?

- Yes!
  - Boosted ability to create deep knowledge about the Sun-Earth system.
  - Enabled timely research by a world-wide community.
  - Novel science is interdisciplinary and benefits from inclusive data-policies.
  - Elimination of proprietary data and proprietary periods has certainly increased the pace of scientific data analysis - PIs can no longer sit on data.
  - Enhanced the competitive environment of scientific discovery.
- Most of the first-look cream-skimming science had been done: times were right for the broadened analyses of complex systems.

# F. Other Comments?

- The evolution must continue on many fronts: particularly at the data analyst's desk.
- Virtual observatories are welcome 'milestones' in the path of the evolution, but not the end point.
- Investigations are moving from case study mode to 'ensemble' study.
- Models need a similar "opening" as data.

- Downside: beware of the 'ignorant' or 'careless' users of data!

# Summary

- During the last solar cycle, the data environment sponsored by NASA's Heliophysics program underwent a significant transition.
- Data sets were opened up, placed on-line, and made available for the research community at large in a timely manner.
- The causes for this shift in paradigm were many: the IT revolution, programmatic guidance, shifts in attitudes about the roles of the PIs, etc.

# Summary (continued)

There were many intended and unintended consequences resulting from this paradigm shift.

- By opening up their data sets, the PIs unintentionally created instantaneous virtual peer groups (the data users) that reviewed data quality on a regular basis.
  - Feedback gained from the broad spectrum of users ultimately improved the data quality and was appreciated by the PIs.
- This transition provided more demand on the data providers to make access easier with appropriate documentation and software tools - creating a positive feedback loop.
- There was a shift in the way the Heliophysics community analyzed their data:
  - from single instrument, point analyses of observed phenomena
  - to a more systematic approach analyzing classes of phenomena by multi-series, multi-instrument investigations by extended groups of analysts.

# Conclusions

- The feedback from the survey was what I expected - for the most part.
- There were some surprises:
  - The respondents gave a much larger role and significance to NASA HQ's guidance and leadership.
  - The role of community peer pressure did not receive that much recognition as a driver for change!
  - The environment is enabling multiplatform, multidisciplinary investigations.
- Thank you to all who participated in the survey.