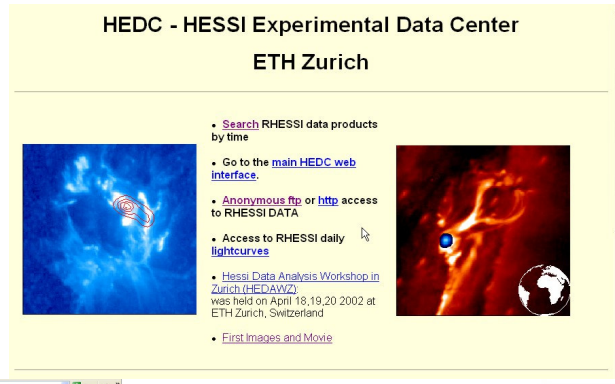
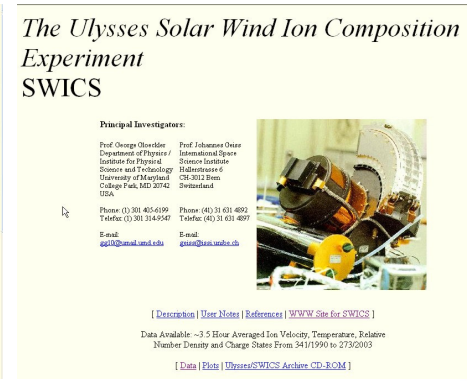
Future Directions in Space Physics Data Management

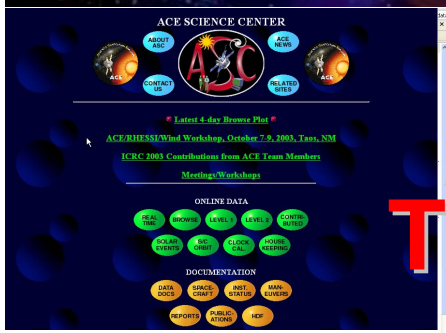
Raymond J. Walker
Todd A. King
Steven P. Joy

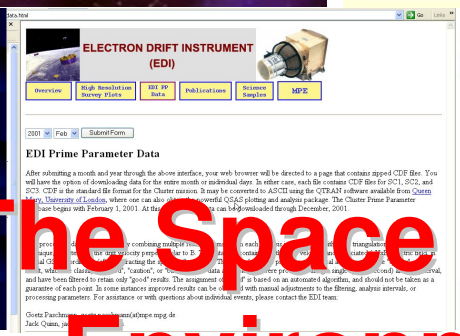
Paper SH42-01
Spring AGU Meeting
New Orleans, LA
May 26, 2005

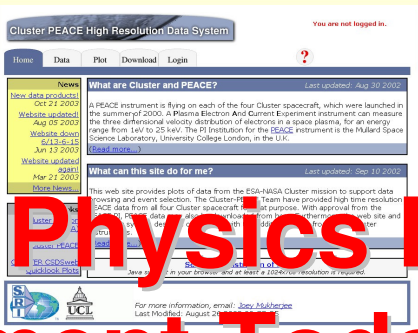
The Space Physics Data Environment Today

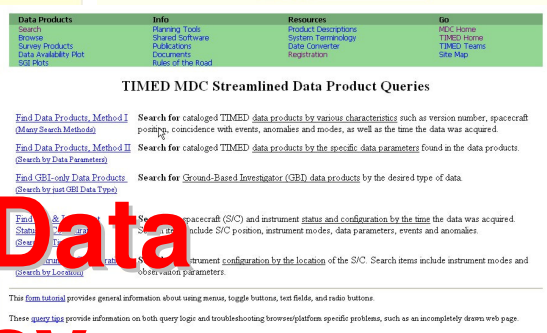


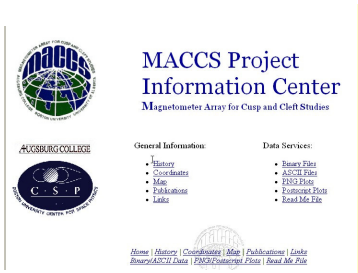


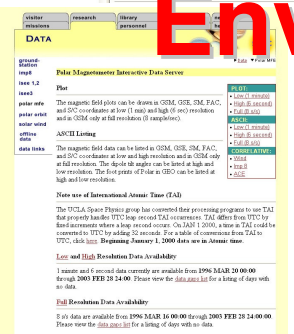


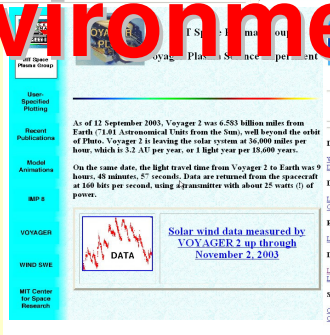


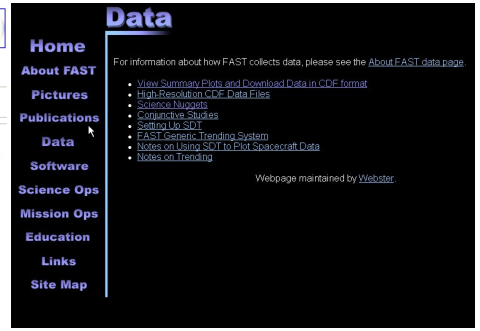


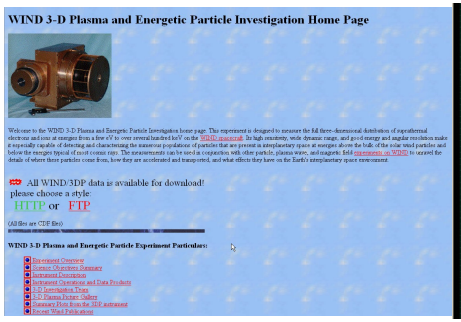


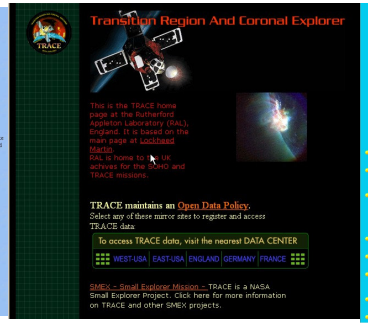


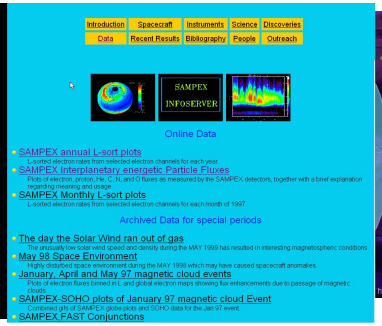


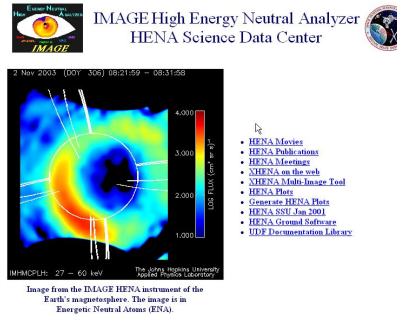












The Data Environment
The Distributed Nature of Space Physics Data

- Space physics data activities are mostly organized by missions.
- Space physics missions mostly have instruments developed and operated by principal investigators rather than facility instruments.
- Although there are a few mission data bases most of the data reside with individual investigators.
- Occasionally data from a given instrument resides at more than one investigator site.
- Space physics requires correlative data from ground based instruments. These data also are widely distributed.

The Data Environment Questions

- How hard is it to obtain and use data from current space physics missions?
  - How easy is it to find the data needed for a given study?
  - If the data were acquired are they available to the scientific community?
  - If the data are available how hard is it to access them?
  - After you have accessed data how hard is it to use them?
  - If you use the data how confident can you be that they are correct?
- Will data be available and useable after the mission ends?
  - Do the mission data repositories have procedures to protect the data from loss?
  - Are the data of archival quality?

The Data Environment
The Answers

Data Access and Use

- How easy is it to find the data needed for a given study?
  - *It is not very easy. There are over 200 sources of space physics data. If you aren't one of the cognizanti it is hard to find the data you need. Multi-spacecraft and multi-instrument studies are very difficult.*

- If the data were acquired are they available to the scientific community?
  - *In general they are available. We are much closer to an open data environment then ever before. Missions need to be encouraged to provide all of the data including the highest resolution data.*

The Data Environment
More Answers

Data Access and Use

- If the data are available how hard is it to access them?
  - *Mixed. Some missions have excellent interfaces and accessing the data is very easy (e.g. ACE, Geotail). Older missions tended to have the worst interfaces.*

- After you have accessed data how hard is it to use them?
  - *Mixed. Many formats are used. Where documentation is good this is not a serious problem. The quality and thoroughness of the documentation is mixed.*

- If you use the data how confident can you be that they are correct?
  - *Mixed. Many missions either do not have data quality programs or left that question blank.*

The Data Environment
More Answers

Data Archiving

- Do the mission data repositories have procedures to protect the data from loss?
  - Yes!

- Are the data of archival quality?
  - No! As noted above the documentation is not uniformly complete. Many data systems have warnings telling users to check with the instrument team before using the data. The instrument team won't always be available. Too many don't get that.

How can we manage this?

SOHO
http://sohowww.nascom.nasa.gov/catalog/
http://sohodata.nascom.nasa.gov/cgi-bin/gui
CDS http://sohodata.nascom.nasa.gov/cgi-bin/gui
CELIAS http://www.cx.unibe.ch/phim/soho/data.html
http://sohodata.nascom.nasa.gov/cgi-bin/gui
http://umtof.umd.edu/pm
http://umtof.umd.edu/pm/figs.html
COSTEP http://sohodata.nascom.nasa.gov/cgi-bin/gui
EIT http://umbra.nascom.nasa.gov/eit/eit-catalog.html
http://sohodata.nascom.nasa.gov/cgi-bin/gui
ERNE http://www.srl.utu.fi/projects/erne/index_english.html
http://sohodata.nascom.nasa.gov/cgi-bin/gui
GOLF http://www.medoc-ias.u-psud.fr/golf/access.html
http://sohodata.nascom.nasa.gov/cgi-bin/gui
LASCO http://lasco-www.nrl.navy.mil/retrieve/products.html
http://lasco-www.nrl.navy.mil/cgi-bin/lwdb/lasco/img_cdrom/form
http://sohodata.nascom.nasa.gov/cgi-bin/gui
http://cdaw.gsfc.nasa.gov/CME_list/index.html
MDI/SOI http://soi.stanford.edu/data/
http://sohodata.nascom.nasa.gov/cgi-bin/gui
SUMER http://www.linmpi.mpg.de/english/projekte/sumer/FILE/SumerEntryPage.html
http://www.linmpi.mpg.de/english/projekte/sumer/text/sff_cat1.html
http://www.linmpi.mpg.de/english/projekte/sumer/text/picture_lnk.html
http://sohodata.nascom.nasa.gov/cgi-bin/gui
SWAN http://sohodata.nascom.nasa.gov/cgi-bin/gui
UVCS http://sohodata.nascom.nasa.gov/cgi-bin/gui
VIRGO http://sohodata.nascom.nasa.gov/cgi-bin/gui

TRACE
http://vestige.lmsal.com/TRACE/

RHESSI
Flight http://rhessidatacenter.ssl.berkeley.edu/
Simulated http://www.hedc.ethz.ch/

ACE
CRIS http://www.srl.caltech.edu/ACE/[illegible]
SIS http://www.srl.caltech.edu/ACE/[illegible]
ULEIS http://www.srl.caltech.edu/ACE/[illegible]
SEPICA http://www.srl.caltech.edu/ACE/[illegible].html
SWIMS http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_SWICS_SWIMS.html
SWICS http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_SWICS_SWIMS.html
EPAM http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_EPAM.html
SWEPAM http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_SWEPAM.html
MAG http://www.srl.caltech.edu/ACE/ASC/level2/lvl2DATA_MAG.html

---

WIND
3-D PLS http://sprg.ssl.berkeley.edu/wind3dp/data/wi/3dp/
ftp://enlil.ssl.berkeley.edu/pub/dat/wi/3dp/
EPACT http://lheawww.gsfc.nasa.gov/docs/gamcosray/lecr/EPACT/epact.html
MFI http://lepmfi.gsfc.nasa.gov/mfi/plots.html
SMS http://space.umd.edu/wind/sms.html
SWE http://web.mit.edu/afs/athena/org/s/space/www/wind.html
TGRS http://lheawww.gsfc.nasa.gov/docs/gamcosray/legr/tgrs/tgrs_home.html
WAVES http://lep694.gsfc.nasa.gov/waves/waves.html

ULYSSES
http://ulysses.jpl.nasa.gov/science/data.html
MAG (VHM/FGM)
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/mag/
http://www.sp.ph.ic.ac.uk/Ulysses/Data/index.html
http://helio.estec.esa.nl/ulysses/ftp/data/vhm_fgm/
http://helio.estec.esa.nl/ulysses/cgi-bin/plots.pl?instrument=VHM_FGM
SWOOPS
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/plasma/swoops/
http://helio.estec.esa.nl/ulysses/ftp/data/swoops/electrons/
http://helio.estec.esa.nl/ulysses/cgi-bin/plots.pl?instrument=SWOOPS-ELECTRONS
http://helio.estec.esa.nl/ulysses/ftp/data/swoops/ions/
http://helio.estec.esa.nl/ulysses/cgi-bin/plots.pl?instrument=SWOOPS-IONS
SWICS
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/plasma/swics/
http://helio2.estec.esa.int/saweb/index.htm
http://helio.estec.esa.nl/Ulysses/ftp/data/swics/
http://helio.estec.esa.nl/Ulysses/cgi-bin/plots.pl?instrument=SWICS
URAP
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/radio/urap/
http://helio.estec.esa.nl/ulysses/archive/urap.html
http://helio.estec.esa.nl/ulysses/cgi-bin/plots.pl?instrument=URAP-SUMMARY
EPAC
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/particle/epac/
http://helio.estec.esa.nl/ulysses/ftp/data/epac/
http://helio.estec.esa.nl/ulysses/cgi-bin/plots.pl?instrument=EPAC
GAS
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/gas/
http://helio.estec.esa.nl/ulysses/archive/gas.html
HISCALE
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/particle/hiscale/
http://sd-www.jhuapl.edu/Ulysses/data.html
http://helio.estec.esa.nl/ulysses/archive/hiscale.html
COSPIN
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/particle/cospin/
http://helio.estec.esa.nl/ulysses/archive/cospin.html
GRB
http://cdaweb.gsfc.nasa.gov/cdaweb/sp_phys/
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/gamma/
http://helio.estec.esa.nl/ulysses/archive/grb.html
DUST
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/dust/
http://helio.estec.esa.nl/ulysses/archive/dust.html
SCE
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/ulysses/radio/sce/
http://helio.estec.esa.nl/ulysses/archive/sce.html
GWE
http://helio.estec.esa.nl/ulysses/archive/gwe.html

VOYAGER
http://voyager.jpl.nasa.gov/science/Voyager_Science_Data.html
ftp://nssdcftp.gsfc.nasa.gov/spacecraft_data/voyager/
[illegible]
[illegible]
[illegible]
http://web.mit.edu/afs/athena/org/s/space/www/voyager/voyager_traj/voyager_traj.html
CRS http://voycrs.gsfc.nasa.gov/heliopause/heliopause.html
LECP http://hurlbut.jhuapl.edu/VOYAGER/index.html
http://voyager.ftecs.com/default.htm
http://space.umd.edu/Voyager/
PWS http://www-pw.physics.uiowa.edu/plasma-wave/voyager/home.html

---

GEOTAIL
http://www.darts.isas.ac.jp/spdb/
LECP http://www.stp.isas.ac.jp/geotail/QL/index.html
MGF http://www.stp.isas.ac.jp/geotail/QL/index.html
CPI http://www-pi.physics.uiowa.edu/www/cpi/
EPIC http://sd-www.jhuapl.edu/Geotail/
PWI http://www.kurasc.kyoto-u.ac.jp/gtlpwi/

CLUSTER
http://cdaweb.gsfc.nasa.gov/
CIS http://cis.cesr.fr:8000/CIS_sw_home-en.htm
EDI http://edi.sr.unh.edu/
EFW http://www.cluster.irfu.se/efw/data/spinfit/index.html
FGM http://www.sp.ph.ic.ac.uk/Cluster/
PEACE http://cluster2.space.swri.edu/
RAPID http://leadbelly.lanl.gov/ccr/level2-data/cgi-bin/get_ccr_digital.cgi
ftp://nis-ftp.lanl.gov user: papco
WBD http://www-pw.physics.uiowa.edu/plasma-wave/istp/cluster/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/cluster/cd/
WHISPER http://cluster2.space.swri.edu/

POLAR
http://pwg.gsfc.nasa.gov/data_products.shtml#instruments
http://cdaweb.gsfc.nasa.gov/
MFE http://www-ssc.igpp.ucla.edu/forms/polar/
http://www-ssc.igpp.ucla.edu/SPDS/polar_orbit/index.html
http://www-ssc.igpp.ucla.edu/forms/polar/corr_data.html
EFI ftp://leproc.gsfc.nasa.gov/pub/polar/efi_kp/
ftp://leproc.gsfc.nasa.gov/pub/polar/efi_kp/sc_potential/
http://lepefi.gsfc.nasa.gov/~polarefi/Online_Data/onlinedata.html
PWI http://www-pw.physics.uiowa.edu/das/po-pwi-list
http://www-pw.physics.uiowa.edu/~jhd/PapCo/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/das/dasInterface.html
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/sfr/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/mca/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/wb/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/restricted/wb/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/lfwr/
http://www-pw.physics.uiowa.edu/plasma-wave/istp/polar/hfwr/
CAMMICE http://leadbelly.lanl.gov/ccr/level2-data/cgi-bin/get_ccr_digital.cgi
http://leadbelly.lanl.gov/ccr/CCR_software.html
http://leadbelly.lanl.gov/ccr/CCR_summary_plots.html
CEPPAD http://leadbelly.lanl.gov/ccr/CCR_summary_plots.html
HYDRA http://www-st.physics.uiowa.edu/www/html/pub_spectra.html
TIMAS ftp://willow.colorado.edu/pub/timas/PAPCO/High_resolution_data.html
ftp://willow.colorado.edu/pub/timas/PAPCO/timas.html
ftp://sierra.spasci.com/DATA/timas/summary/index.html
TIDE http://satyr.msfc.nasa.gov/TIDE/papco.shtml
http://satyr.msfc.nasa.gov/TIDE/software/
http://goewin.gsfc.nasa.gov/cgi-bin/TIDE/summary_plots-dir/summary_gif_walk_2.pl
http://tide.gsfc.nasa.gov/custom_plots.shtml
http://satyr.msfc.nasa.gov/TIDE/plot_orbit_gsml.shtml
UVI http://science.nasa.gov/uvi/LatestImage.htm
http://uvisun.msfc.nasa.gov/UVI/brwseimages.html
http://uvisun.msfc.nasa.gov/UVI/nrtimages.html
PIXIE http://pixie.spasci.com/
http://muir.spasci.com/pixie/homepage/img_dir/graphical_archive.html
http://pixie.spasci.com/pixie/homepage/img_dir/daily/data/
http://muir.spasci.com/pixie/homepage/img_dir/internal/image_of_the_day.html
http://muir.spasci.com/pixie/homepage/movies/current_movie.html
http://muir.spasci.com/pixie/homepage/movies/graphical_archive.html
http://muir.spasci.com/cgi-bin/pix_movie_mkr.html
http://pixie.spasci.com/DynMod/
VIS http://eiger.physics.uiowa.edu/~vis/software/
http://eiger.physics.uiowa.edu/~vis/images/
http://eiger.physics.uiowa.edu/~vis/images/anims/
http://eiger.physics.uiowa.edu/~vis/images/previous_images.html
http://eiger.physics.uiowa.edu/surv[illegible]

[illegible]
[illegible]/sprg.ssl.b[illegible]/welcome.html
[illegible]/fastcgi/sumplotsNetscape/recentplot.pl
EFD http://sprg.ssl.berkeley.edu/htbin/fastcgi/sumplotsNetscape/recentplot.pl
TEAMS http://sprg.ssl.berkeley.edu/htbin/fastcgi/sumplotsNetscape/recentplot.pl
MAG http://sprg.ssl.berkeley.edu/htbin/fastcgi/sumplotsNetscape/recentplot.pl

---

SAMPEX
http://surya.umd.edu/www/sampex.html
HILT http://surya.umd.edu/www/sampex.html
LICA N/A
MAST **N/A**
PET **N/A**
DPU **N/A**

IMAGE
http://150.144.211.77/image/image_main.html
LENA http://lena.gsfc.nasa.gov/
http://150.144.211.77/image/image_main.html
MENA http://pollux.space.swri.edu/mena/
http://150.144.211.77/image/image_main.html
HENA http://sd-www.jhuapl.edu/IMAGE/
http://150.144.211.77/image/image_main.html
RPI http://ulcar.uml.edu/framesRPI.htm
http://150.144.211.77/image/image_main.html
EUV http://euv.lpl.arizona.edu/euv/
http://150.144.211.77/image/image_main.html
FUV http://sprg.ssl.berkeley.edu/image/
http://150.144.211.77/image/image_main.html
CIDP http://150.144.211.77/image/image_main.html
SC http://150.144.211.77/image/image_main.html
OA http://150.144.211.77/image/image_main.html

TIMED
http://www.timed.jhuapl.edu/scripts/mdc_home.pl
GUVI http://guvi.jhuapl.edu/data/dataproducts.shtml
SABER http://saber.larc.nasa.gov/
SEE http://lasp.colorado.edu/see/
TIDI
http://tidi.engin.umich.edu/html/go?access_data/main_data_access.html&menu_data.htm
l

How can we manage this?
Metadata and the Data Model

- Metadata are the glue that binds distributed systems.

- Metadata adheres to an organization defined in a data model.
  - A data model is the keystone of an information system.
  - Scientists (domain experts) define the data model.

- A discipline wide data model for space physics is being developed by the SPASE consortium (paper SH42A-03).

How can we manage this?
The Self-Organizing Data System

The Data System is a collection of repositories

- Archives organized according to a data model
- Repositories are typically distributed

- The first step in developing the self-organized data system is adopting a common data model and dictionary.
  - The SPASE consortium is developing a candidate dictionary.
  - It is important for every stake holder to join or at least become familiar with this effort.

- Repositories document their resources according to the data model by using the shared data dictionary.
  - Resources include data, documentation, software, services, etc.
  - Each resource has a metadata envelope that allows it to be part of a bigger system.
  - Repositories grow resource by resource at their own rate.

How can we manage this?
The Self-Organizing Data System Continued

The Data System self-organizes based on the metadata.

- Repositories are concerned only about resources under their purview.
- The Data System is concerned about how these resources are semantically related.
- All metadata are stored in registries
  - Typically co-located with the repository, but this can be at a remote site.
- The Data System queries all registries to generate a response.
  - A query to the system will always return the most complete answer, even with dynamic (changing) repositories.

How can we manage this?

Connecting the dots

- Semantic technology uses ontologies (formal specifications useful in conceptualization) to discover the relevance of resources in a single domain and between multiple domains.
  - An ontology defines the relationships between objects defined in the data model.

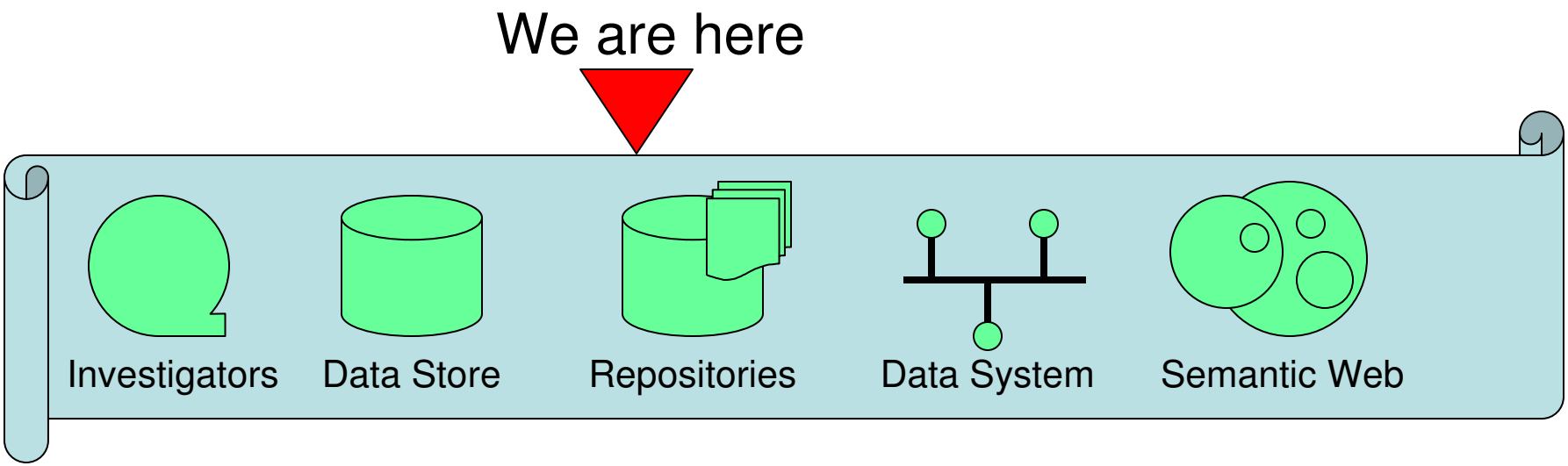
- Formally mapping these relationships gives meaning to the metadata. For example,
  - Are spacecraft and observatories equivalent?
  - Are "different" instruments, like cosmic ray detectors and energetic particle detectors, equivalent?

How can we manage this?
Connecting the dots continued

- Some semantic features can be handled in the data model through abstraction. (You don't always need a semantic layer.) This usually applies only within a single domain.

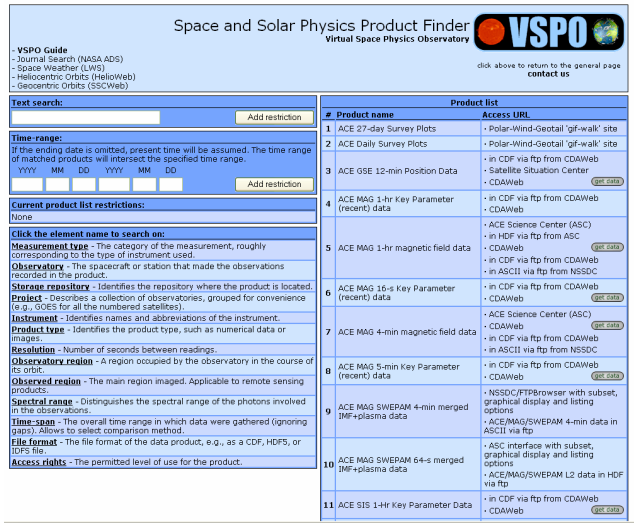
- Domain-to-domain relevance (for example between earth science and space physics) requires semantic technology (to resolve issues of meaning).

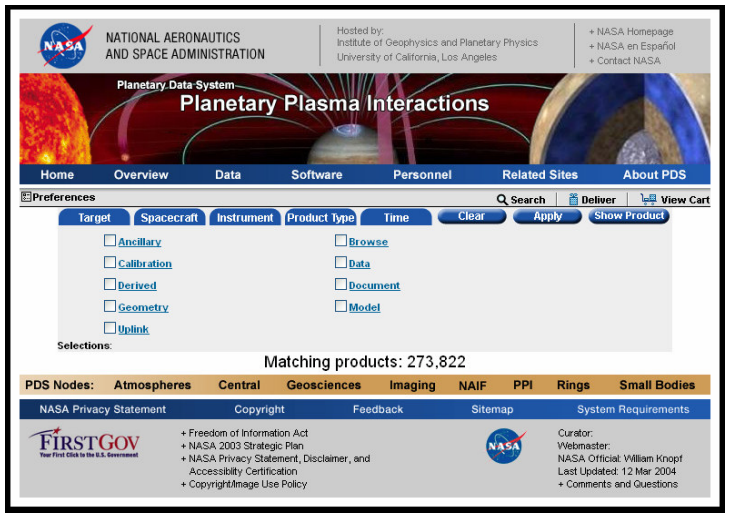
Where are we today?



What can the future be like?

- Going to a single web presence to find any space physics data.
  - Just as Google points you to actual pages, this service will direct you to actual resources.
  - The interface may be a single text box (like Google) or more structured to ensure the proper context for terms.
    - Note: Semantic technology can be used on user input to ascertain context.
  - Resources will be distributed with access transparent to the user.
  - New discoveries will be made by asking new questions.
    - The system will draw inferences using semantics.





This is where we start to talk about the future? Some points to consider.

- Importance of metadata-high data quality and useable data
- Importance of data model- do it by discipline
- SPASE and resident archive programs
- How to you build a data system which is intrinsically distributed?
- What do you handle cross discipline research?
- I would like to at least outline principles that we could follow-