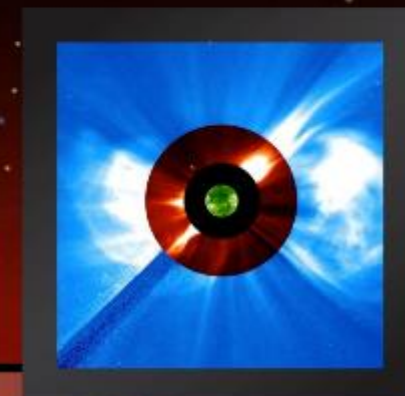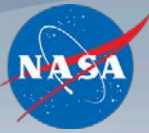**Ad hoc Big Data Task Force**
February 16, 2016

Jeffrey J.E. Hayes
Heliophysics Division
Science Mission Directorate

Why Heliophysics?

Heliophysics is the scientific endeavor to understand 3 fundamental questions…

- What causes the Sun to vary?
- How do the geospace, planetary space environments and the heliosphere respond?
- What are the impacts on humanity?

In order to answer these questions, Heliophysics is implementing a balanced program to achieve 3 goals:

- Explore the physical processes in the space environment from the Sun to the Earth and throughout the solar system;
- Advance our understanding of the connections that link the Sun, the Earth, planetary space environments, and the outer reaches of our solar system;
- Develop the knowledge and capability to detect and predict extreme conditions in space to protect life and society and to safeguard human and robotic explorers beyond Earth.

Heliophysics Mission Portfolio

NASA Heliophysics division seeks to advance the Agency's strategic objectives in heliophysics as well as the science priorities of the Decadal Survey in Heliophysics.

In addition to operating space missions, the Heliophysics portfolio includes basic research and technology development, development and stewardship of national capabilities for conducting space heliophysics, and suborbital investigations (rockets, balloons, and CubeSats). Fundamental to this stewardship is the curation and archiving of NASA's Heliophysics mission data.

Mission investment choices are informed by the Decadal Surveys, other NRC studies, as well as other science community input, particularly advisory committees and peer reviews.
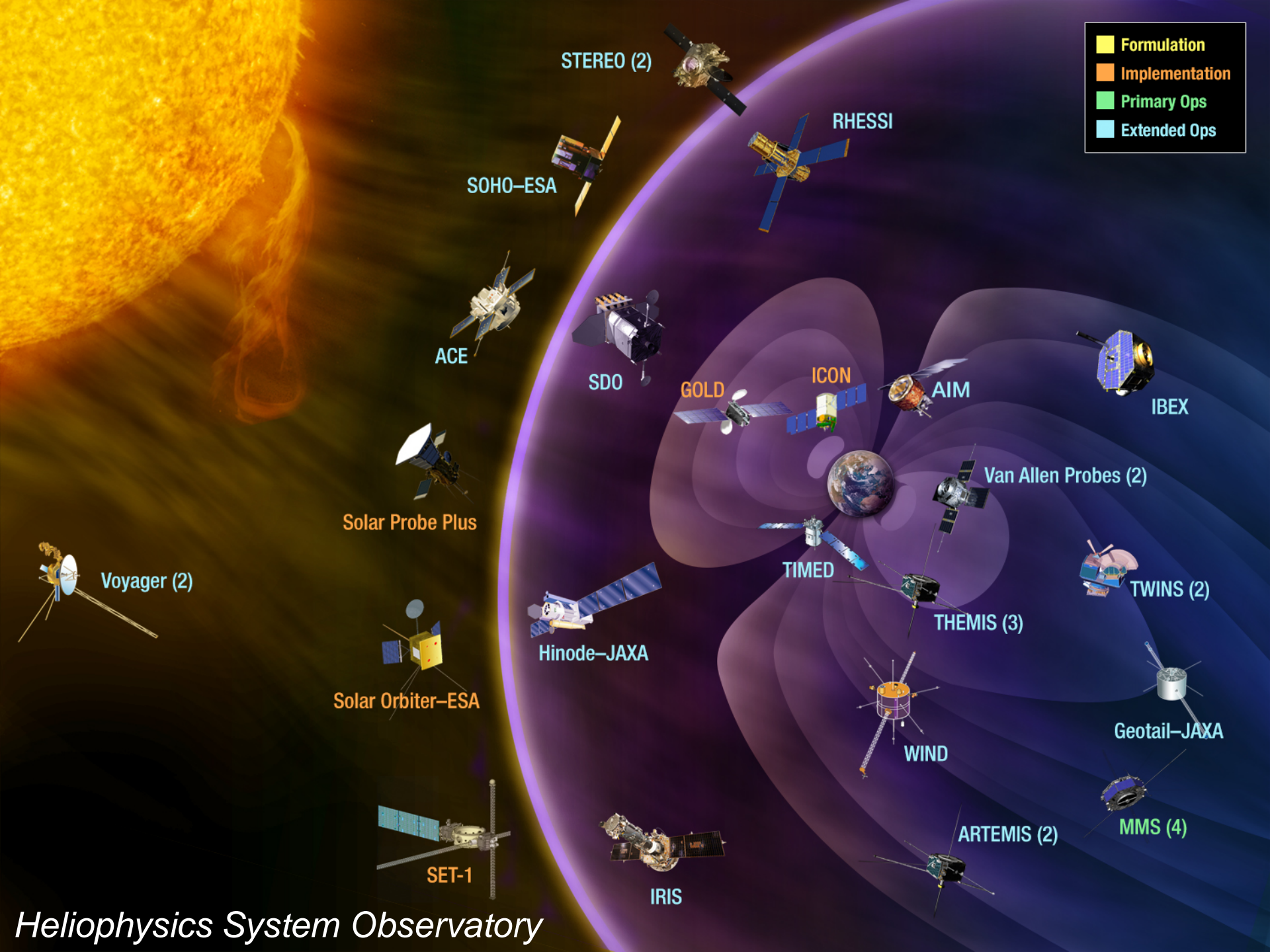
Heliophysics System Observatory

NASA Heliophysics

Please list the missions currently in operational mission phase.

| Mission	Launch	Phase	Extension to –	Type	FY16 Budget ($M)
Geotail	7/24/1992	Extended	12/31/2016	Probe/JAXA	0.233
STEREO	10/25/2006	Extended	9/30/2018	Probe	9.500
THEMIS+Artemis	2/17/2007	Extended	9/30/2018	Explorer	4.600
AIM	4/25/2007	Extended	9/30/2018	Explorer	3.036
Hinode	9/23/2006	Extended	9/30/2018	Probe/JAXA	7.250
ACE	8/27/1997	Extended	9/30/2018	Explorer	3.000
RHESSI	2/5/2002	Extended	9/30/2018	Explorer	1.900
SOHO	12/2/1995	Extended	9/30/2018	Flagship/ESA	2.150
TIMED	12/7/2001	Extended	9/30/2018	Probe	2.670
Voyager 1 + 2	8/20/1977	Extended	9/30/2018	Flagship	5.576
TWINS A + B	6/2006 & 3/2008	Extended	9/30/2018	Explorer/MoO	0.615
IBEX	10/19/2008	Extended	9/30/2018	Explorer	3.400
Wind	11/1/1994	Extended	9/30/2018	Probe	2.150
SDO	2/11/2010	Extended	9/30/2018	Flagsip	9.487
Van Allen Probes	8/30/2012	Extended	9/30/2018	Probe	15.521
IRIS	6/27/2013	Extended	9/30/2018	Explorer	7.715
MMS	3/12/2015	Prime	9/1/2017	Flagship	30.138

NASA Heliophysics

The FY2016 breakout of the Heliophysics budget. Data archiving and curation is part of the Research budget, while some active mission data archiving is in the Operating Missions line.
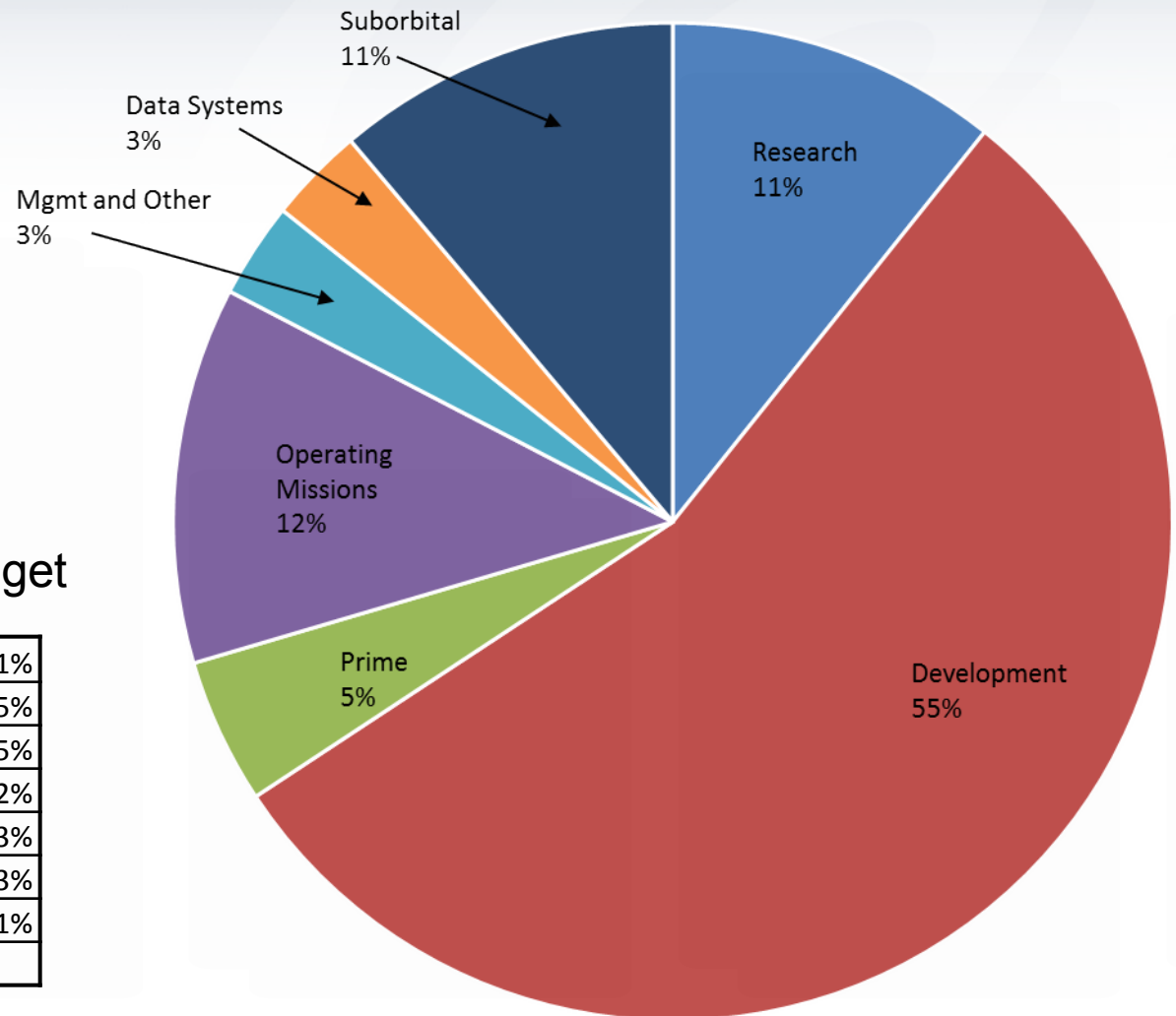
FY2016 Heliophysics Budget

Research	68,658	11%
Development	352,027	55%
Prime(MMS)	30,138	5%
Operating Missions	78,170	12%
Management and Other	19,811	3%
Data Systems	19,890	3%
Suborbital	71,420	11%
Total	**640,114**	

Heliophysics Mission Portfolio
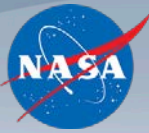
Heliophysics missions reflect the primary classes of SMD missions

- Strategic Missions
  - Initiated by NASA generally in response to recommendations in the Decadal Survey
  - NASA-led strategic Heliophysics missions are generally in the large or medium mission class
  - NASA also initiates strategic partnerships with other space agencies, generally resulting in a NASA contribution to a partner-led mission
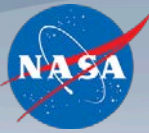
- PI-led competed missions
  - Initiated by a PI-led team in the form of a Heliophysics Explorers proposal to NASA, either for a full mission or a mission of opportunity (MO)
  - Heliophysics Explorers full mission classes are small (SMEX) and medium (MIDEX) size
  - Mission of opportunity classes included contributions to a partner-led mission, small complete missions for the cost of a MO, and suborbital-class missions

Examples	Full Mission	Contribution
Strategic	MMS, SPP	Solar Orbiter
Competed	ICON, GOLD	CINDI, TWINS

Senior Reviews and extending missions: its all about the data

NASA carries out Senior Reviews to assess whether a mission extension should be extended past its original planned life. These reviews take a number of factors into account:

Scientific merit;
Promise of future impact and Productivity;
Progress towards previously agreed-to Prioritized Mission Objectives (PMOs);
Impact of past scientific results;
Accessibility, usability, and utility of data;
Spacecraft and instrument health & safety;
Productivity and vitality of the science team;
Level and quality of the stewardship of the asset;
Effectiveness of communications to the general public.
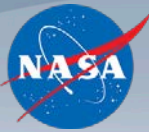
Extended Missions – the source of data

Heliophysics is a field that encompasses a number of sub-discipline areas which are all intimately interconnected:

- Solar physics: the processes on the Sun and its extended outer atmosphere (photosphere, chromoshere, and corona) and the kinetics driving these observed phenomena inside the star. The Sun is mildly variable with a periodicity of ~22 years;

- Solar wind and the heliosphere: the region between the outer corona and the planets encompassing the entire solar system out to ~121 AU (as measured by Voyager 1) containing all manner and energies of particles, waves, and fields;

- The magnetosphere: an extremely complex and dynamic region around the Earth caused by the existence of the Earth's magnetic field which responds dramatically with solar wind variations;

- The ionosphere, thermosphere, and mesosphere (ITM): the upper portions of the Earth's atmosphere, which also respond to solar wind variations as part of the connection of the Sun to the Earth.

The Sun's 22-year cycle means that phenomena in differing places within this complex, coupled system will look differently depending on where, and when in the solar cycle one observes them.

Extend Missions Costs

To place the opportunity costs for the extended operations of the Heliophysics suite of missions in context, in FY16, the entire Heliophysics budget is $650M.

The entire portfolio of all operational Heliophysics missions is $99.9M (this includes MMS, which is in its prime mission). If one excludes MMS, then the total for all the extended operating missions is $78.8M. This represents 12% of the entire Heliophysics budget for this fiscal year, or an <u>average</u> investment of $4.9M per mission. Indeed if one were to breakdown the cost per satellite (because some missions have multiple spacecraft involved) the average cost per satellite is even lower, at $2.9M.

All active missions are archiving and curating data within the costs discussed above.   In addition, the missions are encouraged to start talking to the two Heliophysics active archives sooner rather than later: the **Solar Data Analysis Center (SDAC)**, and the **Space Physics Data Facility (SPDF)**. These represent a total investment of ~$3.3M per year and are both at GSFC. In addition, the Heliophysics Division competes a small program to capture old, and generate new, higher-level data products. This Research Opportunities in Space and Earth Sciences (ROSES) element is ~$1M per year with 1 or 2 year awards.

What scientists want to do

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

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

What scientists want a data system to do

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

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!

Heliophysics Data Environment Current Status

• The Virtual (discipline) Observatories (VxOs) and the Heliophysics Data Centers have, over recent years*, produced a comprehensive set of dataset descriptions* in standard terms, thus allowing *ready access to Heliophysics data* (e.g., through SPDF's HP Data Portal).   The *Virtual Observatory* paradigm is "**universal access through standards**."

• Heliophysics Final Archives (SDAC, SPDF) have been systematically *working with new missions* (Van Allen Probes, BARREL, MMS, Solar Probe, IRIS) to assure that *data will be easily available for the short and long term*. A comprehensive set of NASA mission data (apart from SDO) is now served from the Final Archives.

• *Web services facilitate direct access* to most Heliophysics data from applications such as IDL, and from others' services (EU AMDA; data/model services; TPLOT).

• We have *established standards for data formats* (FITS, CDF, and NetCDF), and for *metadata* (SPASE), and these are stable but responsive to community needs.

• We have *restored a large fraction of the data from older missions* (ISEE-1, 2, 3; DE 1, 2; Helios 1, 2; AMPTE; FAST; Yohkoh; etc.), and these data are accessible in standard formats from the HP Archives.   We are systematically examining the *NSSDC data archives* and restoring datasets of scientific interest.

• *Data are no longer dropped when missions end*, and even the "Resident Archives" that were needed for this are becoming nearly obsolete.
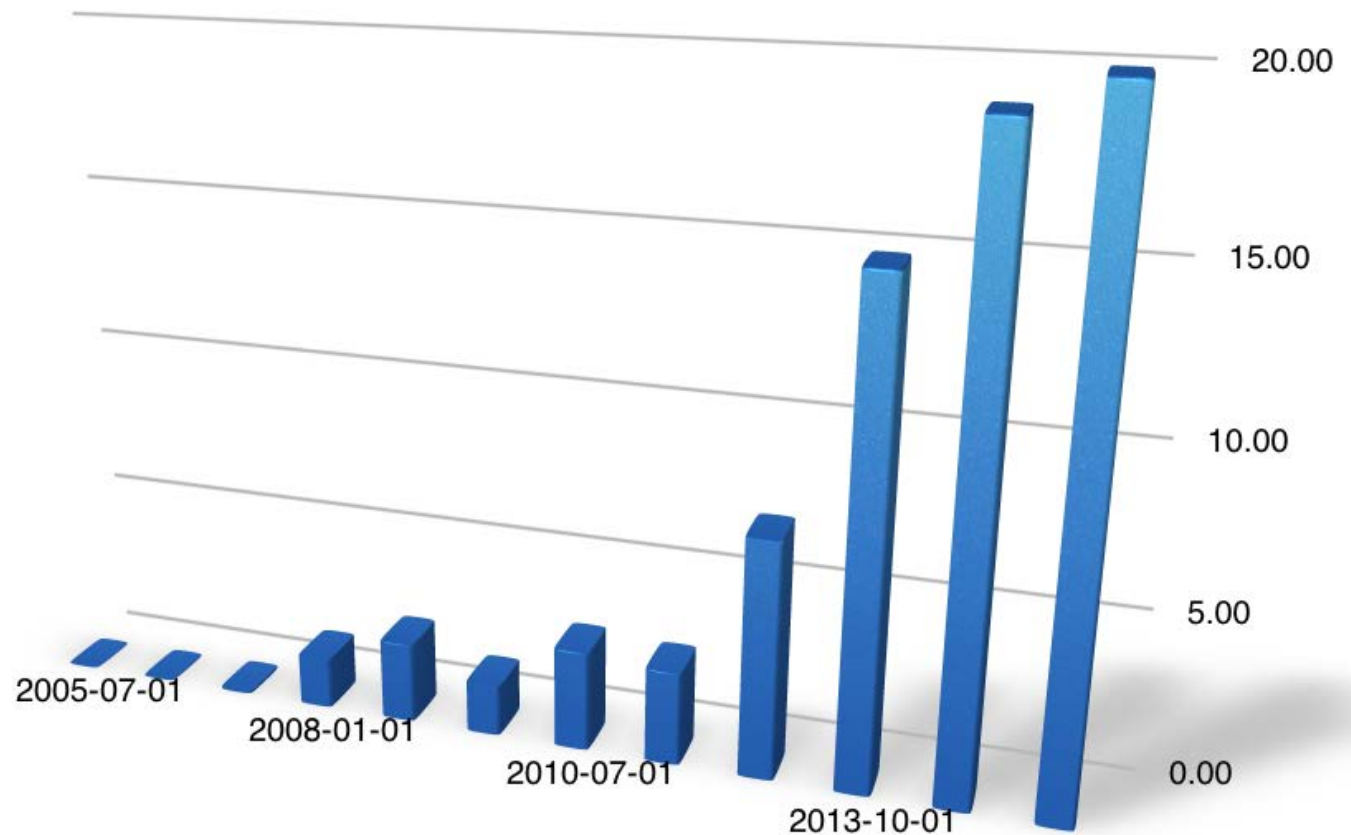
Trends:
Estimated monthly data downloads from VSO (Tbyte)
(quarterly averages; some data missing)
Does *not* include European mirrors

Before mid-2010, figures include all data sources

After mid-2010, only SDO AIA and HMI data for simplicity

Volume of AIA downloaded from Lockheed Web services (*e.g.* cut-out service) is similar (~ 20 Tbyte/month).

What the Solar Data Analysis Center (SDAC) currently serves

- STEREO and *Hinode* data sets include all instrument data.

- SOHO and TRACE includes most data.

- 1.04PB total archived; 9 missions + 1 rocket.

- 125M files and growing.

- ~325TB per month served.

- VSO acts as discovery portal and allows ground-based data to be accessed along with NASA data.
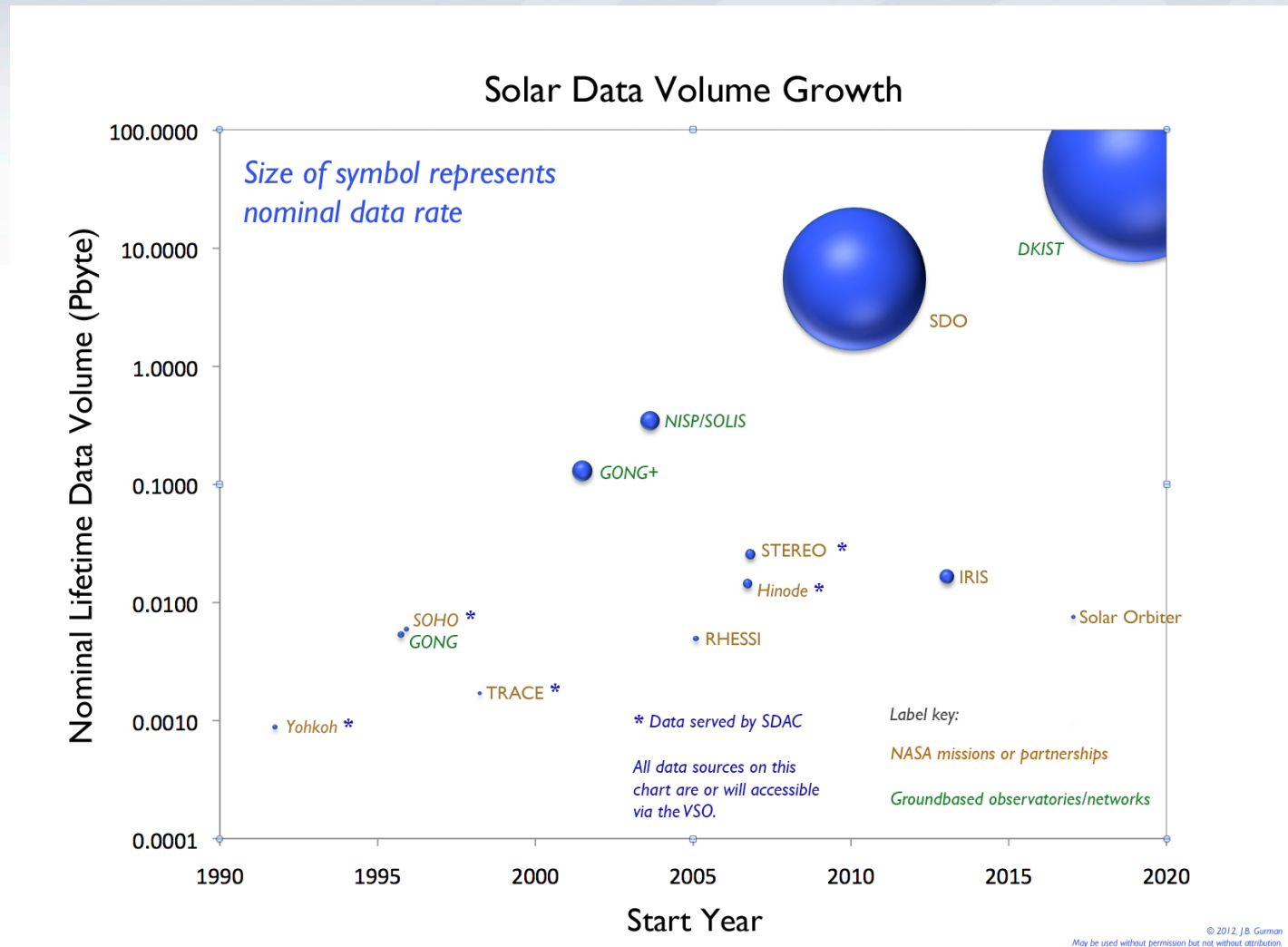
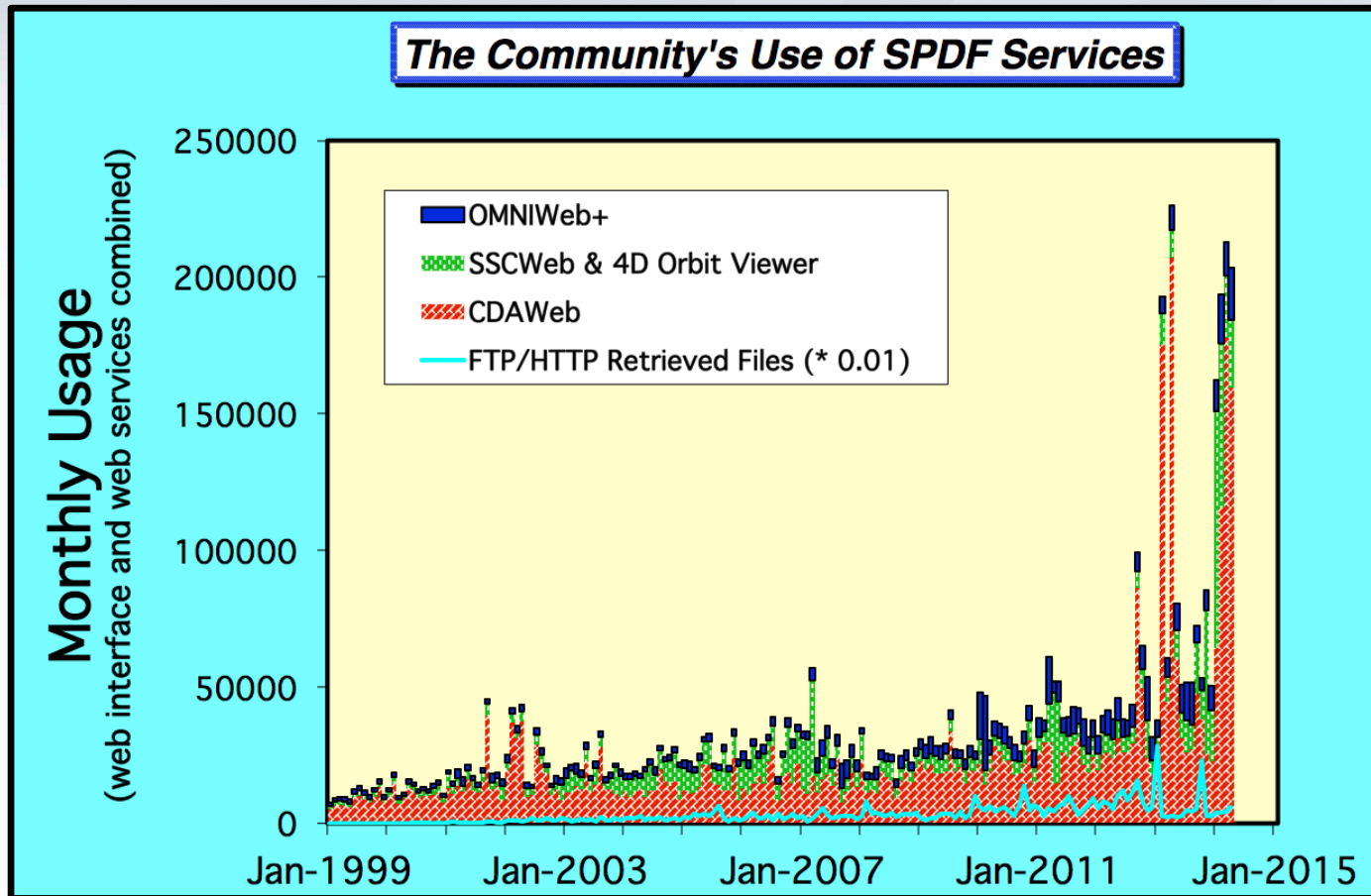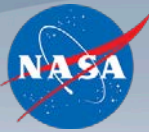| Data Source	Volume (Tbyte)	Active?
STEREO	71.0	Yes
Hinode	12.7	Yes
SOHO	14.0	Yes
TRACE	24.0	No
SDO cache	800.0+	Yes
other	15.7	No

Solar Project Data Volume Growth

Both lifetime data volume and rate continue to grow

There is and will continue to be experience in the solar groundbased community as well as the NASA-supported community

Thus it makes sense to share experience and best practices between the two communities
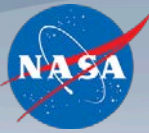

Solar Data Volume Growth

Size of symbol represents nominal data rate

DKIST

SDO

NISP/SOLIS

GONG+

STEREO *

Hinode *

IRIS

SOHO *
GONG

RHESSI

Solar Orbiter

TRACE *

Yohkoh *

* Data served by SDAC

All data sources on this chart are or will accessible via the VSO.

Label key:

NASA missions or partnerships

Groundbased observatories/networks

Nominal Lifetime Data Volume (Pbyte) vs Start Year

© 2012, J.B. Gurman
May be used without permission but not without attribution.

SPDF Use & Acknowledgements

The Community's Use of SPDF Services

91TB of holdings as of Feb 2016: MMS to be made public 3/1/2016.
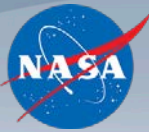23 current spacecraft: ~130 past. Monthly ~6TB are served from and
~730k files.

Growth in SPDF Archive Over Time

Archived TBytes at SPDF
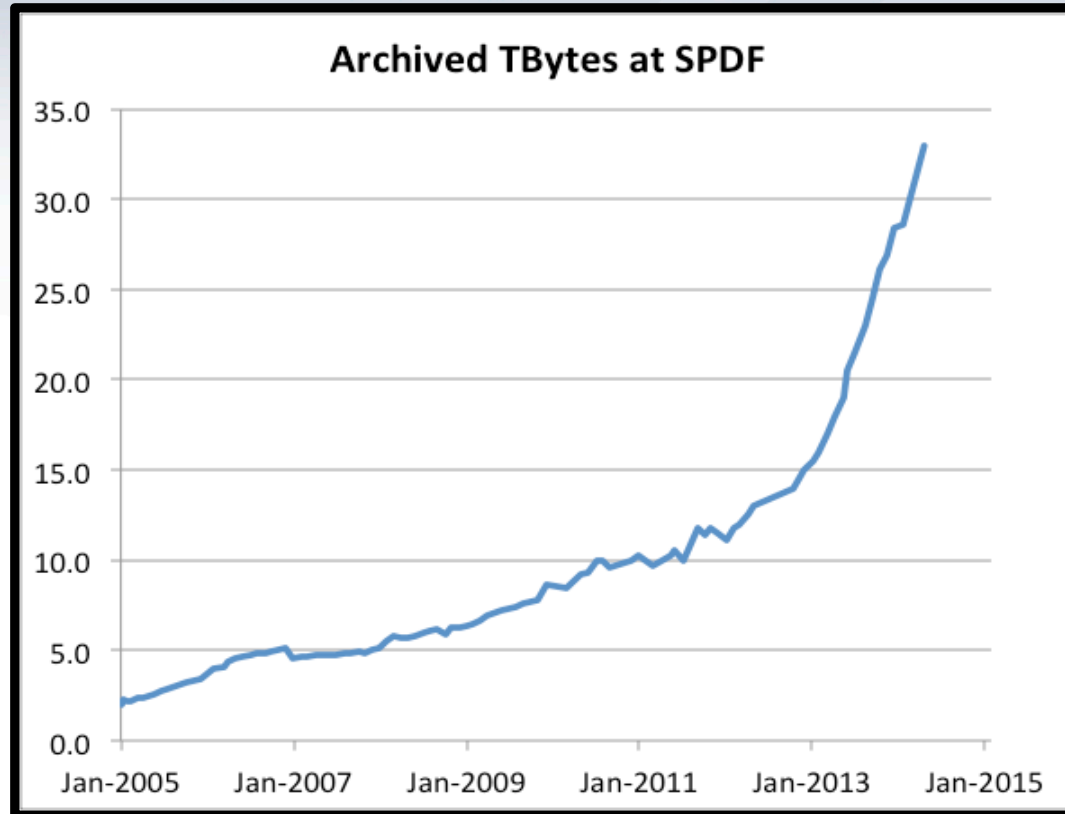
- Dominant data flow is Van Allen Probes and starting 3/1/2016, MMS.
  - Significant current data from Cluster WBD instrument, as well as more comprehensive archival data are coming on line *e.g.* from TIMED.

The Heliophysics Data Policy

The Heliophysics Division issued a Data Policy in 2007 (revised in 2009) emphasizing NASA's open data policy and the need to archive and curate data in standard formats, which are now explicitly specified. This policy involves the producers of data as responsible partners in insuring usability and accessibility of NASA taxpayer-funded science endeavors to the scientific community and the public.

New missions are following this policy and are delivering data as expected; VxOs and Final Archives are involved in the process.

Current missions are improving their data, documentation, and services; most are in good shape.

Senior Reviews and Mission Archive Plans are leading to better archives.

Data are moving into Active Final Archives, even from the most recent missions, and are being served and kept safe.

An Inventory and Registry of all Heliophysics data is being completed and has an active interface (Heliophysics Data Portal) that will deliver or point directly to data.

Legacy datasets are being improved, archived, and served.

Plans are moving forward for uniform access to heliophysics data.
  – Standards, standards, standards…

The Heliophysics Data Policy is working.
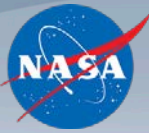
Future Challenges & Vision

- Metadata production and use (Core Data Enhancements focus)
  - Definitive inventory/registry: referential (Digital Object Identifiers?) and discovery uses.
  - Uniform data access for all products.
  - Seamless flow from mission archives through to final archives.
- <u>Format standards</u> (e.g., CDF-A; also a NetCDF standard? FITS is well established.)
  - Adoption of standards in calls for mission proposals (the community now agrees!).
- Large data volumes
  - How to use the data: Pattern recognition; data mining.
  - How to keep the data available and safe post-mission (SDO is most significant case for now).
- Model-data comparisons and insights
  - Seamless integration of model output with data streams.
  - Data assimilation; true space weather capabilities.
  - Data volume questions, as above.
- Plans for follow-on to "VxOs": Consolidate access; Develop Value-Added Services as independent projects that use the infrastructure.
  - Provide one "reference implementation" "HVSO" for general data access (NOT 9 routes to data).
  - Focus on providing metadata and links to generic access methods, especially for new missions via the Core Data Enhancements.
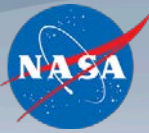
Acronyms

- ACE – Advanced Composition Explorer
- AIM – Aerometry of Ice
- AMDA – an online tool for data analysis sponsored by the Centre de Donnees de la Physique des Plasmas
- AMPTE – Active Magnetospheric Trace Explorers
- Artemis – (Goddess of the Moon: 2 THEMIS probes in orbit there))
- BARREL – Balloon Array for RBSP Relativistic Electron Losses
- CDF – Computable Document Format
- CINDI – Coupled Ion Neutral Dynamic Investigation
- DKIST – Daniel K. Inoe Solar Telescope
- FITS – Flexible Image Transport System
- GOLD – Global-scale Observations of the Limb and Disk
- GONG – Global Oscillation Network Group
- GONG+ – see above (as augmented)
- HVSO – Heliophysics Virtual Solar Observatory
- IBEX – Interstellar Boundary Explorer
- ICON – Ionospheric Connection Explorer
- IDL – Interactive Data Language
- ISEE – International Sun-Earth Explorer
- IRIS – Interface Region Imaging Spectrograph
- MMS – Magnetospheric MultiScale mission
- NetCDF – Network Computable Document Format
- NISP/SOLIS – National Solar Observatory Integrated Synoptic Program/Synoptic Optical Long-term Investigations of the Sun
- RA – Resident Archive

Acronyms (Cont'd)

- RHESSI – Reuven Ramaty High Enegry Solar Spectroscopic Imager
- SDO – Solar Dynamic Observatory
- SET-1 – Space Environment Testbeds
- SOHO – Solar and Heliospheric Observatory
- SPASE – Space Physics Archive Search and Extract
- STEREO – Solar TErrestrial RElations Observatory
- THEMIS – Time History of Events and Macroscale Interactions during Substorms
- TIMED – Thermospshere Ionosphere Mesosphere Energetics and Dynamics
- TB/PT – terabyte/petabyte
- TPLOT – interactive online plotting tool
- TRACE – Transition Region & Coronal Explorer
- TWINS – Two Wide-angle Imaging Neutral-atom Spectrometers
- VSO – Virtual Solar Observatory
- VxO – Virtual (discipline) Observatory