



NASA's Earth Science Data Systems: A "Bit of History" and Observations

Presented at
Cyberinfrastructure for Environmental Observations,
Analysis, and Forecasting:
A Cyberinformatics Forum

H. K. "Rama" Ramapriyan
Earth Science Data and Information System Project
Rama.Ramapriyan@nasa.gov
May 7, 2008
(Charts revised July 1, 2008)



Contents

- **NASA's Earth Science Data Systems**
- **Core and Community Systems**
- **Data Systems Context**
- **The Alphabet Soup**
 - **EOSDIS/ ESIPs/ NewDISS/ SEEDS/ ESDSWG/ REASoN/ ACCESS/ MEaSURES**
- **EOSDIS – Key Metrics**
- **Data Systems' Evolution (70s to present, and future)**
- **Observations**
- **Vision 2015**



NASA's Earth Science Data Systems

- **“Study Earth from space to advance scientific understanding and meet societal needs” -- *2006 NASA Strategic Plan***
- **NASA's Earth Science Data Systems directly support this objective by providing end-to-end capabilities to deliver data and information products to users**



Core and Community Capabilities - Definition

- **‘Core’ data system elements reflect NASA’s responsibility for managing Earth science satellite mission data characterized by the continuity of research, access, and usability.**
- **The core comprises all the hardware, software, physical infrastructure, and intellectual capital NASA recognizes as necessary for performing its tasks in Earth science data system management.**
- **‘Community’ elements are those pieces or capabilities developed and deployed largely outside the NASA core elements and are characterized by their ‘evolvability’ and innovation.**

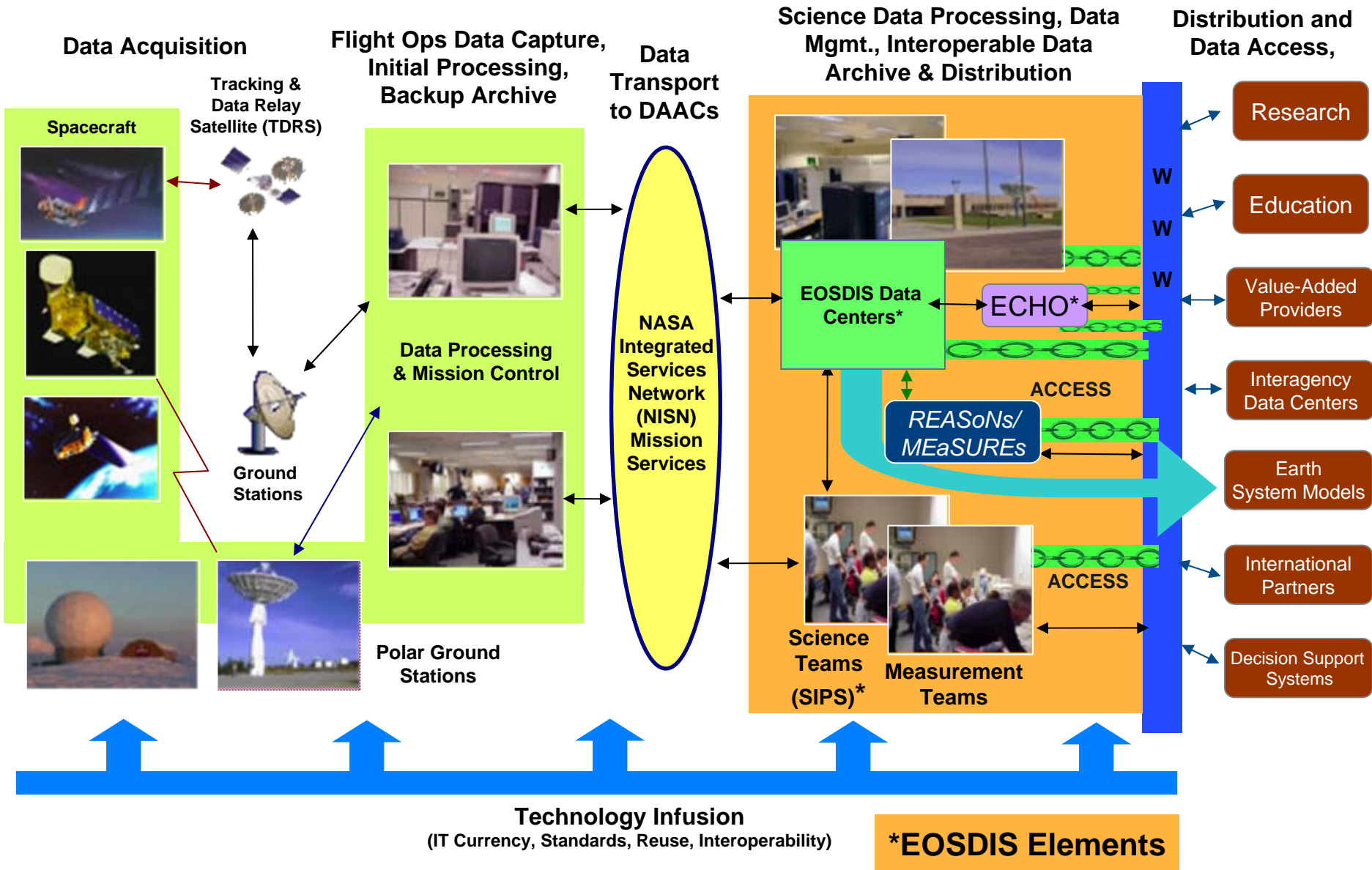


Core and Community Capabilities - Characteristics

| CORE | COMMUNITY |
|--|---|
| Projects Subject to Programmatic Review | Projects Competitively Selected |
| Substantive NASA Oversight | 'Light Touch' Oversight w/Significant Community Involvement |
| Tight Integration of Data System Tools, Services and Functions | Community-based Tools and Services Loosely-Coupled |
| Employ Well Established Information Technologies | Employ 'Edgy' or Emerging Technologies |

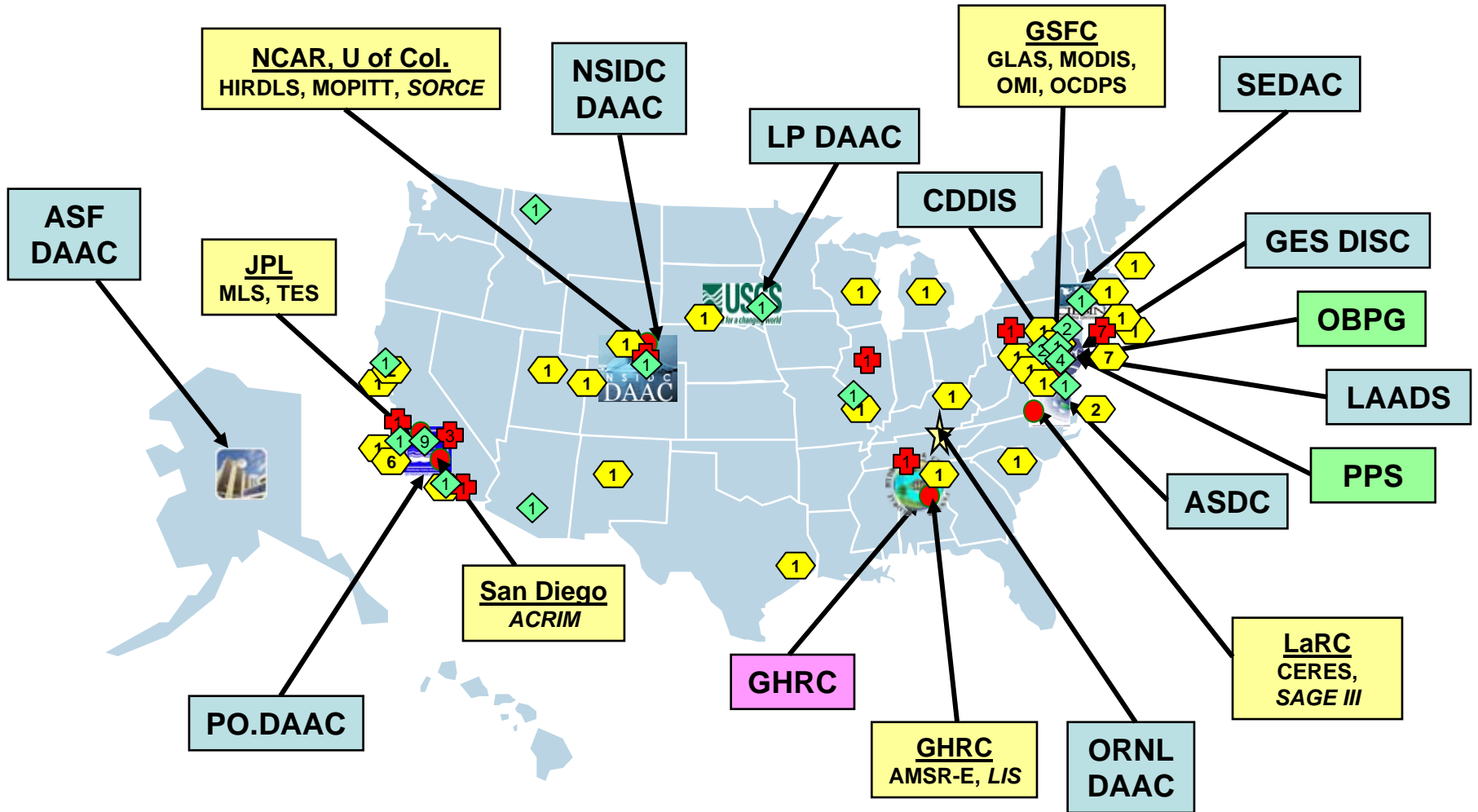


Earth Science Data Systems Context





NASA Earth Science Data Systems (Core and Community)



KEY

EOSDIS Data Centers

Related Data Providers

Measurement-based Systems

Science Investigator-led Processing Systems (SIPSS)

41

REASoN

29

MEaSUREs

17

ACCESS



Alphabet Soup (1 of 3)

- **EOSDIS – Earth Observing System Data and Information System**
 - Operating since 1994, starting with “Version 0” managing heritage (pre-EOS) data at Distributed Active Archive Centers (DAACs) and making them interoperate
 - Now managing all of EOS mission data and derived standard data products
- **ESIPs – Earth Science Information Partners**
 - Federation initially supported (1998-2003) by NASA in response to 1995 NRC review recommendation; experiment in “self governance”
 - Currently over 90 partners – government funded as well as commercial
- **NewDISS – New Data and Information Systems and Services –** study initiated in 1998 by NASA HQ with an internal/external group led by Martha Maiden – “Draft Version 1.0” report dated February 2002
- **SEEDS – Strategic Evolution of Earth Science Enterprise (ESE) Data Systems – Formulation Study** conducted by a NASA GSFC team “chartered to work with the Earth science user and data provider communities to generate approaches and plans for future ESE data and information systems” – Final Report July 2003



Alphabet Soup (2 of 3)

- **ESDSWG – Earth Science Data System Working Groups – Recommended by SEEDS study and approved by NASA as a mechanism for community input on data system issues**
 - **Four WGs are active: Metrics Planning and Reporting, Software Reuse, Standards Processes, Technology Infusion**
 - **MPARWG:** Review and recommend program-level performance metrics and collection tools that measure how well each data activity supports the NASA Earth Science Division's (ESD's) research, application and education programs
 - **Reuse WG:** NASA ESD spends a significant amount of resources developing software components that may have value to other NASA programs in terms of functionality and/or applicability. The Software Reuse Working Group is chartered to oversee the process that will maximize the reuse potential of such components ...
 - **SPG:** Welcomes (and seeks out) submissions of potential standards that would be of value to the NASA Earth Science community. These standards are evaluated and can eventually be endorsed as ESDS standards.
 - **TIWG:** Enable NASA's Earth Science Enterprise to reach its research, application, and education goals more quickly and cost effectively through widespread adoption of key emerging information technologies.



Alphabet Soup (3 of 3)

- **REASoN – Research, Education, and Applications Solutions Network Program – 42 projects initiated in 2003/ 2004 through NASA’s Cooperative Agreement Notice**
 - Provide data products, information systems and services capabilities, and/or advanced data systems technologies integrated into the project, to address strategic needs in Earth science research, applications, and education.
- **ACCESS – Advancing Collaborative Connections for Earth System Science Program – 17 projects initiated in 2005/ 2006**
 - Enhance and improve existing components of the distributed and heterogeneous data and information systems infrastructure
- **MEaSUREs – Making Earth System data records for Use in Research Environments – 29 projects initiated in 2007/ 2008 (Some completed REASoN Projects are continuing under this program)**
 - Create Earth System Data Records (ESDRs), including Climate Data Records
 - An ESDR is defined as a unified and coherent set of observations of a given parameter of the Earth system, which is optimized to meet specific requirements in addressing science questions.
 - Such records are critical to understanding Earth System processes, to assessing variability, long-term trends, and change in the Earth System, and to provide input and validation means to modeling efforts.



Core Capability – EOSDIS

- **NASA's Earth Observing System Data and Information System (EOSDIS) is a petabyte-scale archive of environmental data that supports global climate change research**
- **EOSDIS provides for**
 - Data ingest
 - Data processing
 - Data distribution
 - Archive management



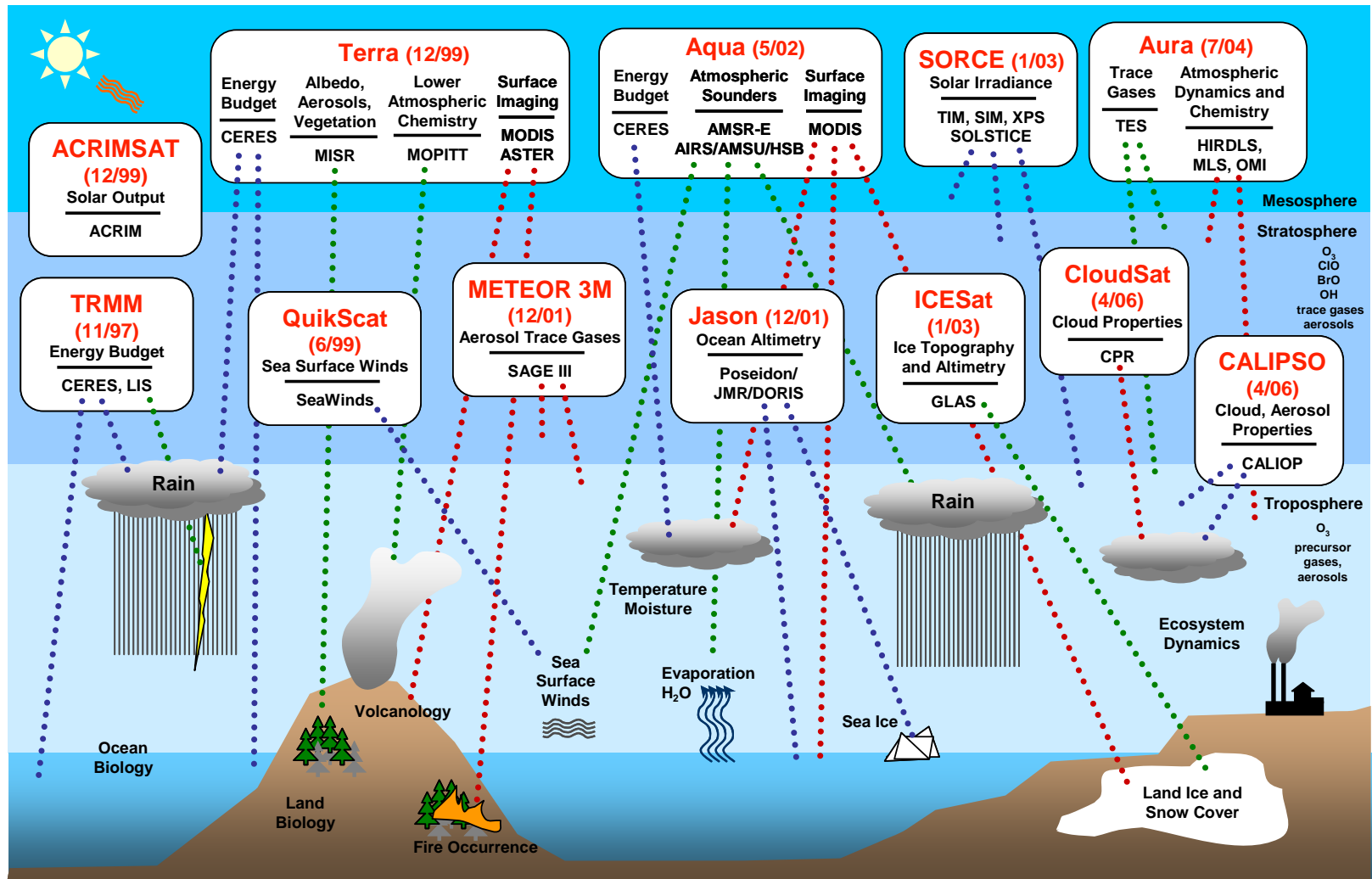
This MODIS image shows the wide sediment plume of the Yangtze River as it empties into the East China Sea.

Credit: Jacques Descloitres, MODIS Land Science Team

Image Date: 09-16-2000



EOSDIS Manages Data For All 24 EOS Measurements

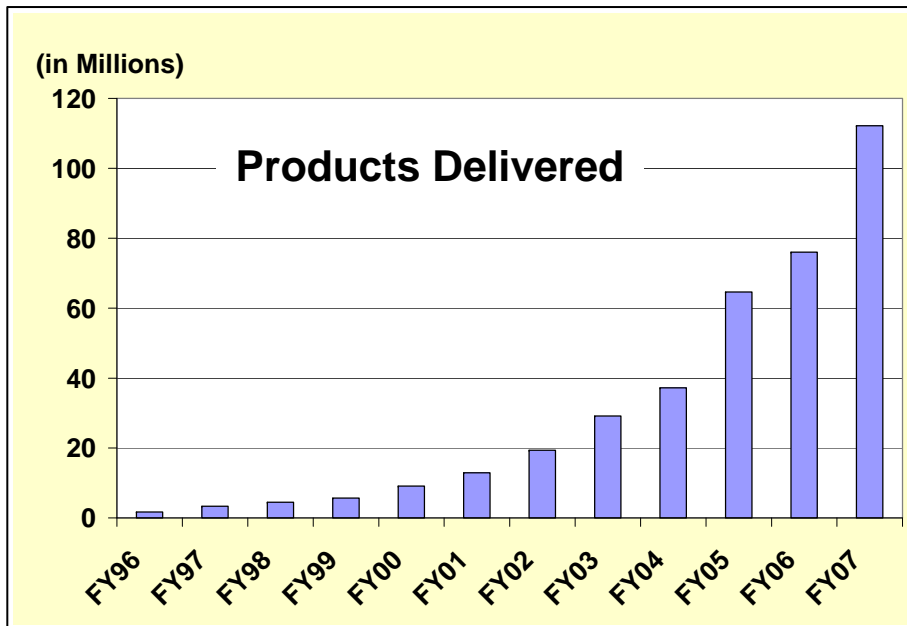




EOSDIS Key Metrics

| EOSDIS Metrics (Oct 1, 06 to Sept 30, 07) | |
|---|------------|
| Unique Data Products | >2700 |
| Distinct Users at Data Centers | ~3.0M |
| Daily Archive Growth | 3.2 TB/day |
| Total Archive Volume | 4.9 PB |
| End User Distribution Products | >100M |
| End User Daily Distribution Volume | 4.2 TB/day |

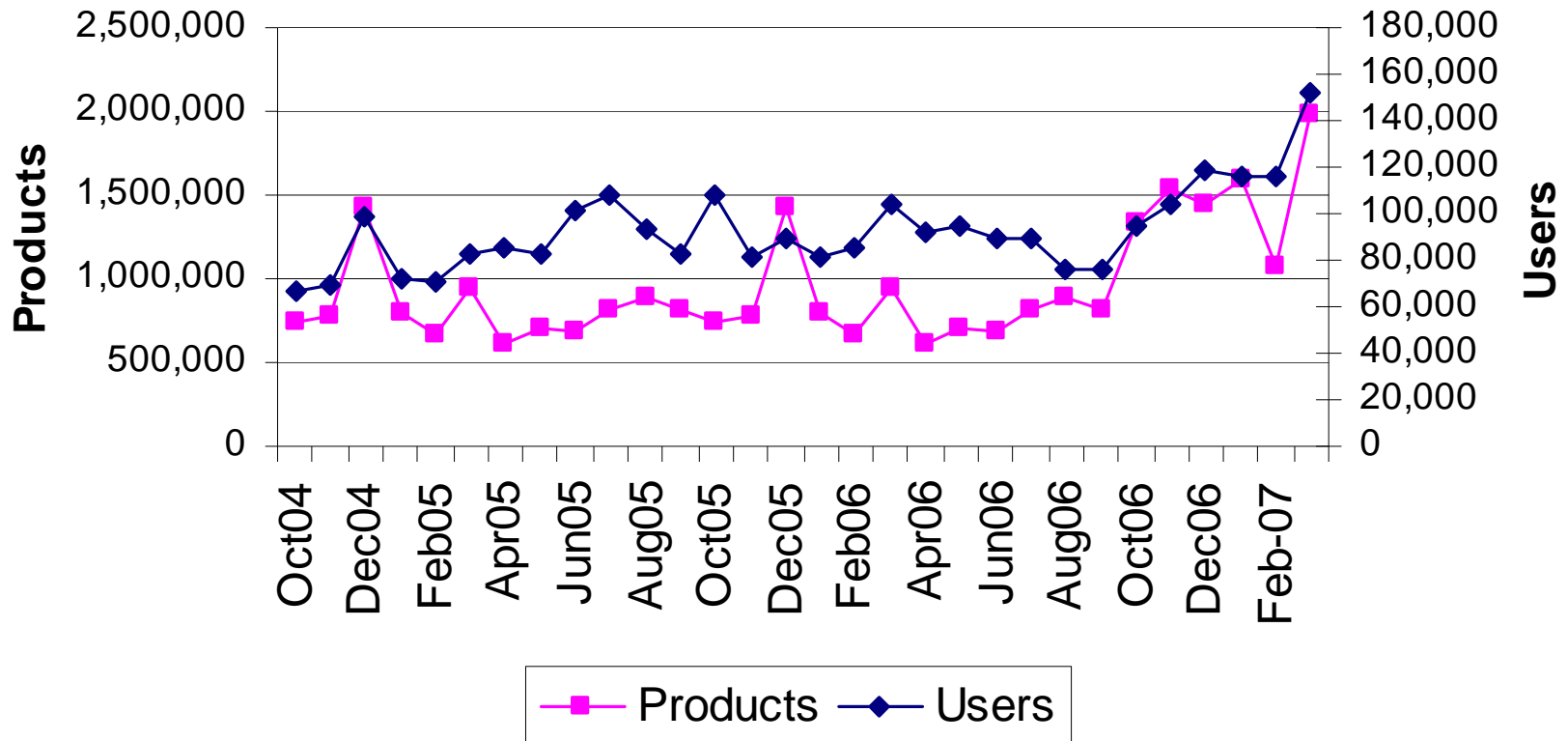
| ESDIS Project Supports | | |
|-------------------------|-----------------------------|----|
| Science System Elements | Data Centers | 11 |
| | SIPS | 14 |
| Interfaces | Interface Control Documents | 41 |
| Partnerships | US | 8 |
| | International | 18 |
| Missions | Science Data Processing | 7 |
| | Archiving and Distribution | 51 |
| | Instruments Supported | 75 |





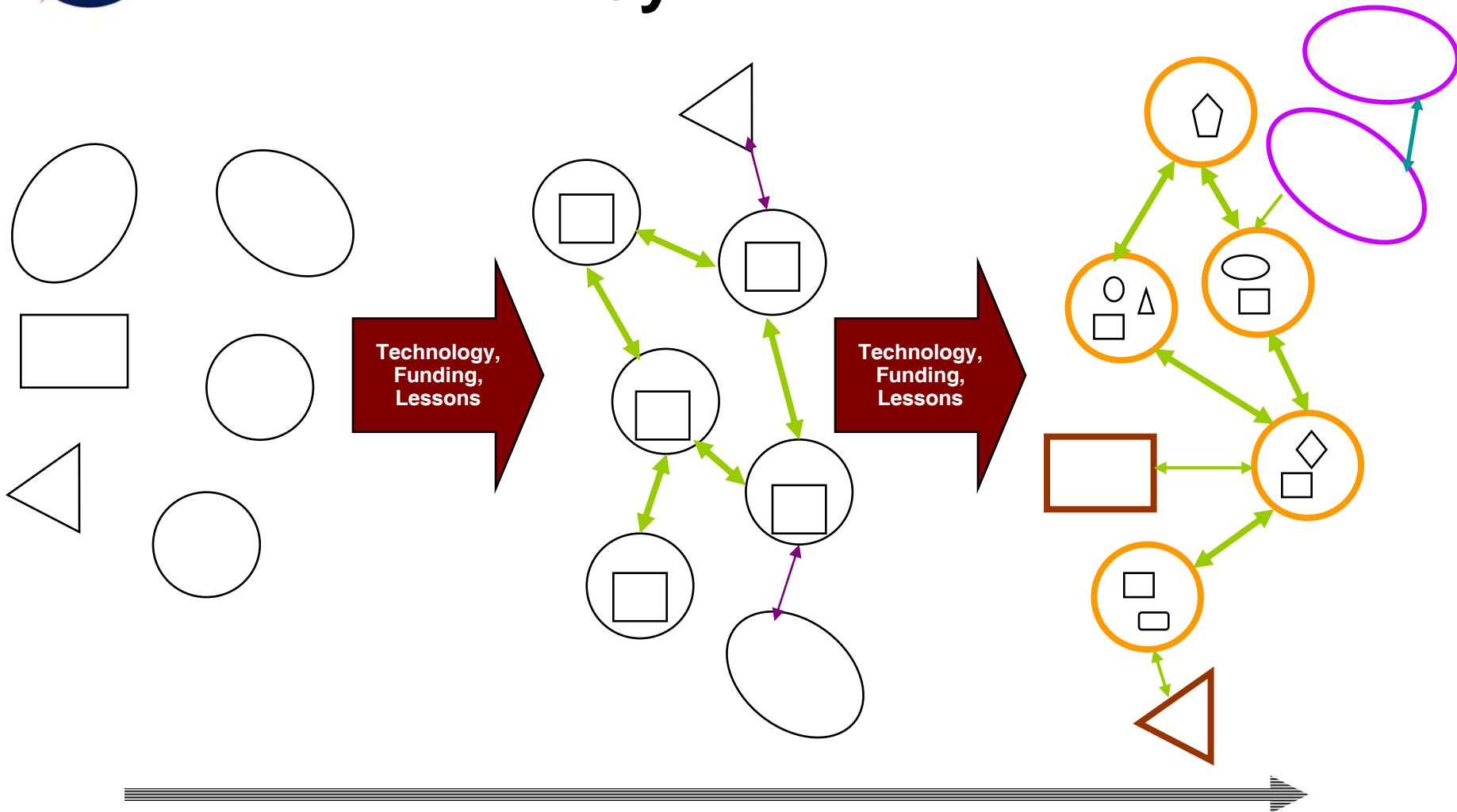
REASoN Metrics

REASoN Projects - Monthly Distinct Users and Products Provided, Oct 2004 - Mar 2007





Data Systems' Evolution



Heterogeneous;
non-interoperable
(<1990)

"Strong" core; some
interoperable,
heterogeneous
components (mid-1990s)

"Simplified" core; greater
autonomy for interoperable,
heterogeneous components
(late 1990s – 2000s)



Evolution of Data System Features

- Discipline and mission specific data systems
- Only community-specific standards
- Only directory functions to point and link users to various data systems
- Multi-site access and data inter-use cumbersome

Technology,
Funding,
Lessons

- Improved access to heritage data
- Cross-system search and order access via interoperability data model
- Also site-specific interfaces
- Common distribution format (HDF); other formats also supported

Technology,
Funding,
Lessons

<1990

mid-1990s



Evolution of Data System Features

Technology,
Funding,
Lessons

- Support for high data volumes and ambitious performance requirements
- Integrated core infrastructure plus loosely coupled elements*
- Common data model
 - Automated metadata creation and ingest
 - No need for cross-site metadata translation
 - FGDC standards compliance
- Expanded set of software tools and services
- Flexible options for supporting or interoperating with external data sources

*mix of EOSDIS Core System, SIPs, Data Center developed capabilities, community capabilities

Technology,
Funding,
Lessons

- Coexistence of heterogeneous, distributed data providers/information partners
- Minimal set of core standards
- Support for community-specific standards
- Reusable software components
- Exploitation of Service Oriented Architecture
- On-line archives and cross-system service invocation
- Ease of innovation and technology infusion

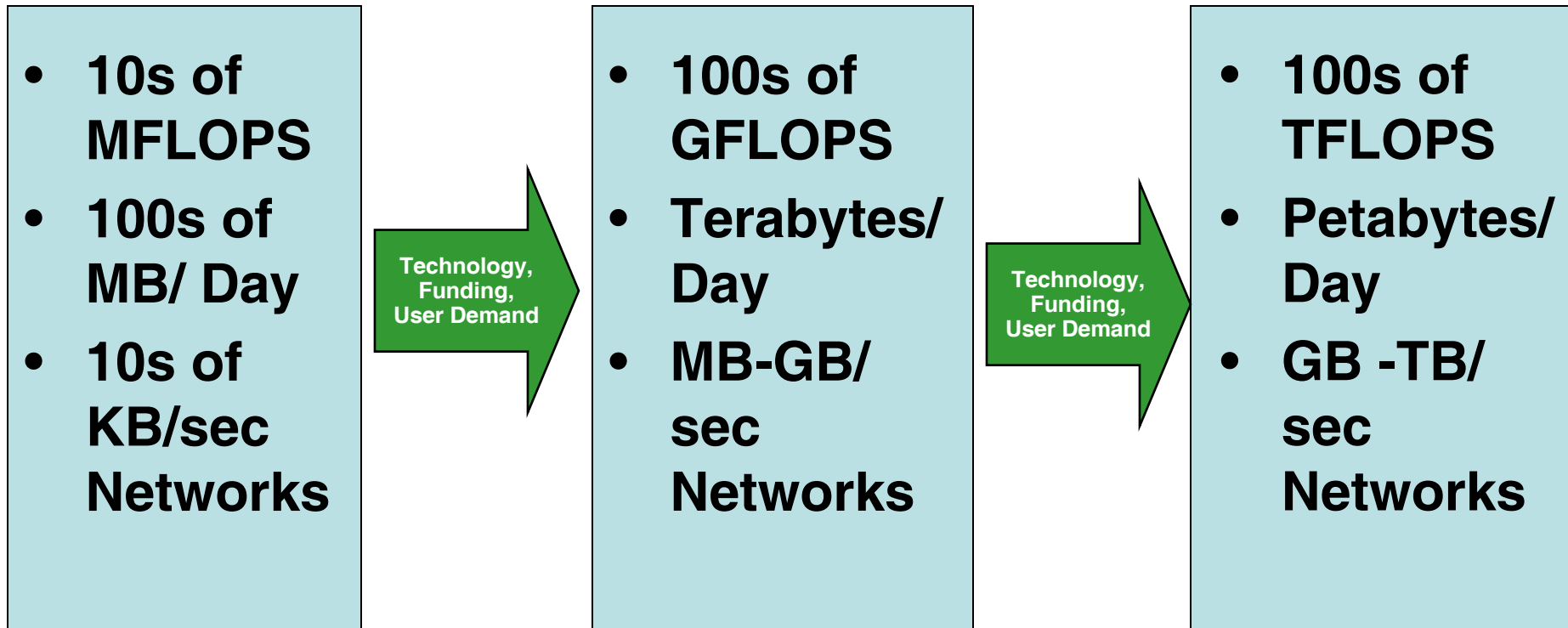
Technology,
Funding,
Lessons

Late 90s to present

Near-Future



Evolution of Capacity Needs



Past

Present

Future



Best Practices (1 of 4)

- Open Data Policy
 - NASA provides open access to data with no period of exclusive access
 - Most of the data are provided at no charge to any requesting user
- Both Core and Community Capabilities are essential to meet NASA's Earth Science program objectives
 - Core capabilities are needed for long-term stability and dependable capture, processing, and archiving of data and distribution of data to a broad and diverse communities of users, including value-added service providers
 - Community capabilities provide innovative, new scientific products as well as a path to technology infusion
 - **NASA currently has four Earth Science Data System Working Groups (ESDSWG) – see <http://esdswg.gsfc.nasa.gov/>**
 - Standards Processes Group
 - Technology Infusion Working Group
 - Reuse Working Group
 - Metrics Planning and Reporting Working Group
 - **Working groups provide community-vetted recommendations to NASA to consider implementation**
 - **These recommendations as well as those from EOSDIS Data Centers, annual user feedback through surveys and at community conferences, interagency and international discussions influence NASA's programmatic direction**
 - **NASA needs to strengthen its effort in facilitating technology infusion from community to core systems**



Best Practices (2 of 4)

- **Loosely coupled, heterogeneous systems can work together (important “existence proof” for GEOSS)**
 - Early development of EOSDIS (so-called Version 0) involved making heterogeneous systems interoperate in the “pre-WWW” era
 - Successful, with well-defined interfaces and a “thin” translation layer to spread queries to multiple databases and gather responses to present to users (“one-stop shopping”)
- **Complex development of EOSDIS Core System (ECS) with “strongly coupled” components proved to be difficult**
 - Eventually successful after reducing scope and allocating most of processing to Science Investigator-led Processing Systems
 - Version 0 Information Management System (IMS) was adopted for one-stop shopping across data centers
 - Managing standards and interfaces was key to success
 - Thorough interface tests and end-to-end testing was critical
- **Community evolution of standards works better than top-down approach**
 - Essential to provide flexibility to accommodate multiple standards and software tools to facilitate data use



Best Practices (3 of 4)

- **One size does not fit all**
 - **Scientific disciplines have different ways of looking at the data and different vocabularies.**
 - **Need flexibility and tools to handle other data and metadata formats**
 - **Need some consistency to facilitate search and access across datasets**
 - **Enable/Facilitate development of different interfaces to support different communities**



Best Practices (4 of 4)

- **Data Systems must evolve over time**
 - In early 2005, NASA embarked on an EOSDIS Evolution Study
 - Addressed multi-faceted goals/issues:
 - Manage archive volume growth
 - Improve response and data access
 - Reduce recurring costs of operations and sustaining engineering
 - Update aging systems and components
 - Move towards more distributed environment
 - A vision for the 2015 timeframe was developed by the EOSDIS Elements Evolution Study Team to guide conduct of study (see end of presentation for Vision 2015)
 - It is critical to manage transitions of an operational system that serves large numbers of users
 - Transitions are made incrementally
 - Each transition involves testing by interfacing systems' staff, and certification by affected users (or representatives)



EOSDIS Evolution 2015 Vision Tenets

| Vision Tenet | Vision 2015 Goals* |
|--|---|
| Archive Management | <ul style="list-style-type: none"> ▪ NASA will ensure safe stewardship of the data through its lifetime. ▪ The EOS archive holdings are regularly peer reviewed for scientific merit. |
| EOS Data Interoperability | <ul style="list-style-type: none"> ▪ Multiple data and metadata streams can be seamlessly combined. ▪ Research and value added communities use EOS data interoperably with other relevant data and systems. ▪ Processing and data are mobile. |
| Future Data Access and Processing | <ul style="list-style-type: none"> ▪ Data access latency is no longer an impediment. ▪ Physical location of data storage is irrelevant. ▪ Finding data is based on common search engines. ▪ Services invoked by machine-machine interfaces. ▪ Custom processing provides only the data needed, the way needed. ▪ Open interfaces and best practice standard protocols universally employed. |
| Data Pedigree | <ul style="list-style-type: none"> ▪ Mechanisms to collect and preserve the pedigree of derived data products are readily available. |
| Cost Control | <ul style="list-style-type: none"> ▪ Data systems evolve into components that allow a fine-grained control over cost drivers. |
| User Community Support | <ul style="list-style-type: none"> ▪ Expert knowledge is readily accessible to enable researchers to understand and use the data. ▪ Community feedback directly to those responsible for a given system element. |
| IT Currency | <ul style="list-style-type: none"> ▪ Access to all EOS data through services at least as rich as any contemporary science information system. |

*Developed by EOSDIS Elements Evolution Study Team - 2005



Conclusions

- **NASA has significantly improved its Earth Science Data Systems over the last two decades**
- **Open data policy and inexpensive (or free) availability of data has promoted data usage by broad research and applications communities**
- **Flexibility, accommodation of diversity, evolvability, responsiveness to community feedback are key to success**



Acknowledgements

- **Materials for this presentation have evolved over many years of the author's work with NASA's Earth Science Data Systems and contain contributions from several people in the Earth Science Data and Information System Project at NASA Headquarters and the Goddard Space Flight Center**
- **The author is particularly grateful for the most recent inputs/comments from Martha Maiden and Frank Lindsay (NASA HQ); Beth Weinstein, Lola Olsen and Christopher Lynnes (NASA GSFC); Yonsook Enloe (SGT, Inc.); and Paul Davis (U of MD)**
- **These charts are revised from the version presented at the May 5, 2008 cyberinfrastructure meeting (see title page). A condensed version of this material was presented by Martha Maiden at the AGU Joint Assembly, Spring 2008 in Ft. Lauderdale, FL**
- **Any opinions expressed here are those of the author and do not necessarily imply official NASA policy**