Evolution of Information Management at the GSFC Earth Sciences (GES) Data and Information Services Center (DISC): 2006–2007

Steven Kempler, Christopher Lynnes, Bruce Vollmer, Gary Alcott, and Stephen Berrick

Abstract—Increasingly sophisticated National Aeronautics and Space Administration (NASA) Earth science missions have driven their associated data and data management systems from providing simple point-to-point archiving and retrieval to performing user-responsive distributed multisensor information extraction. To fully maximize the use of remote-sensor-generated Earth science data, NASA recognized the need for data systems that provide data access and manipulation capabilities responsive to research brought forth by advancing scientific analysis and the need to maximize the use and usability of the data. The decision by NASA to purposely evolve the Earth Observing System Data and Information System (EOSDIS) at the Goddard Space Flight Center (GSFC) Earth Sciences (GES) Data and Information Services Center (DISC) and other information management facilities was timely and appropriate. The GES DISC evolution was focused on replacing the EOSDIS Core System (ECS) by reusing the in-house developed disk-based Simple, Scalable, Script-based Science Product Archive (S4PA) data management system and migrating data to the disk archives. Transition was completed in December 2007.

Index Terms—Data management, Earth science data systems, information management (IM), information technology, online archives, remote sensing.

I. INTRODUCTION

IN 2005, National Aeronautics and Space Administration (NASA) Earth science information management evolution shaping forces lined up to permit an evolution acceleration, which was implemented during 2006–2007, that has greatly improved the way NASA Earth science data centers archive, distribute, and manage data and advanced information services. The decision by NASA to purposely evolve the Earth Observing System (EOS) Data and Information System (EOSDIS) was timely and appropriate. Up to this point, NASA's investment in EOS has yielded dozens of missions, greatly enhancing our understanding of the planet's land, oceans, and atmosphere [1]. Missions were formulated, and science investigations were selected around six interdisciplinary Science Focus Areas, based on NASA's Earth science strategic goal, "Study Earth from space to advance scientific understanding and meet societal needs," and subsequent Earth science questions [2]. Increasingly sophisticated NASA Earth science missions have driven their associated data and data management systems from providing simple point-to-point archiving and retrieval to performing user-responsive distributed multisensor information extraction. To maximize the use of remote-sensor-generated Earth science data, NASA recognized the need for, and needed investment in, data systems that provide data access and manipulation capabilities responsive to research brought forth by advancing scientific analysis, as well as the need to maximize the use and usability of the data, thus providing more scientists with NASA resources for research analysis. The employment of responsive data management systems and information technologies to facilitate science research has been accomplished through the development of the EOSDIS and associated Distributed Active Archive Centers (DAACs), Principal Investigator (PI) processing systems, Earth Science Information Partners (ESIPs) [3], and various NASA Research Announcements (NRAs), seeking the best and most innovative ideas for advancing NASA Earth science data systems and technologies, on behalf of furthering science.

Obviously, in step with science research, science data and information systems will always evolve. Nowadays, at the doorstep of formulating future missions recommended by the National Research Council's Decadal Study [4], or the like, data and information systems must continue to be responsive to new missions. The objectives of the 2006–2007 evolution of EOSDIS were to "increase end-to-end data system efficiency while decreasing operations costs, increase data interoperability and usability by the science research, application, and modeling communities, improve data access and processing, and ensure safe stewardship" [5]. The information management system evolution that occurred in 2006–2007 will benefit science with the ability to integrate more adaptable data manipulation capabilities. These integrations will occur in response to science needs. "The steps we take today to evolve EOSDIS ... should make it more agile and adaptable to change" [5]. This paper provides a brief history of NASA Earth science data and information management evolution, followed by the shaping forces that ultimately drove the evolution of 2006–2007. With the stage set, a description of the evolution of 2006–2007, which occurred at NASA's Goddard Space Flight Center (GSFC) Earth Sciences (GES) Data and Information Services Center (DISC), one of NASA's DAACs, is given.

II. BRIEF HISTORY OF NASA EARTH SCIENCE DATA AND INFORMATION MANAGEMENT

A. Middle Ages

"In the 1970s and 1980s NASA's Earth science data were managed using two distinct approaches. With one approach the data were held by PIs or at specialized data systems. The other approach used a central data system for processing, archiving and distributing the data" [3]. In both cases, PIs and their teams became the primary sources of the data. In the former approach, PI teams acquired and processed the raw data, usually holding on to it for long periods of time before offering it to the broader research community. Whereas this was done often to ensure data validation, sometimes, it was to ensure that the science team can perform first science on the data. The result of the first approach was limited data access and science. The second approach, using a central data system for data operations, was not significantly better. For example, in the case of the Upper Atmospheric Research Satellite (UARS), "PIs were required to deliver their product generation software to a Central Data Handling Facility (CDHF) where data were processed and archived" [3]. However, even in this case, CDHF data access was limited to the UARS PI teams for a period of 2 years. In general, no data standards existed, broad distribution was virtually unheard of, and interoperability was not possible.

B. Renaissance

Starting in the mid-1980s, with the growth of Internet communications, allowing the movement of relatively large amounts of data, and the advent of desktop computing, providing data analysis computing power to nonlocal (to the archive) researchers, the need was realized for open data archive and distribution sites. Three pilot programs were born: The Pilot Climate Data System (PCDS), later named the NASA Climate Data System (NCDS), at the NASA GSFC, managed "a large collection of climate-related data of interest to the research community," providing "uniform data catalogs, inventories, access methods, graphical displays, and statistical calculations" for selected data sets [7]. The Pilot Ocean Data System (PODS) at the Jet Propulsion Laboratory (JPL) was developed to "investigate techniques for archiving and distributing ocean data obtained from space . . . (and) permit researchers to extract and use data rapidly and conveniently" [8]. The Pilot Land Data System (PLDS), was developed at GSFC "to store satellite-, aircraft-, and ground-acquired data; to remotely access this data and information about the data; and to transmit the data to distant geographical locations" [9].

To further facilitate the public availability of satellite-generated data, NASA initiated the implementation of EOSDIS in the early 1990s, "meant to collect, process, distribute, and archive the large amounts of data to be generated by the EOS program" [10]. NASA's Earth science data would be more publicly available and more conducive to interdisciplinary science. Specific Standard Data products derived from EOS instruments, utilizing standard data formats and corresponding documentation, would be archived at one of eight, specialized by Earth science discipline, DAACs. Each DAAC would have access to science and data experts and provide data and user services for their discipline. One system, EOSDIS Core System (ECS)

would be developed and deployed at all DAACs to perform "core" common data ingestion, archiving, and distribution, and user services. (Note: It turned out that ECS was deployed at just four DAACs.) The "theory" was to standardize data systems. In addition, each DAAC would develop an in-house data management system, called Version 0 (V0), which would archive and distribute existing data sets (e.g., subsume the "Pilot" data sets), and be a prototype for the ECS, until ECS was deployed.

Concurrently, on the data user side, NASA instituted Regional Earth Science Applications Centers (RESACs), selected to innovatively "use NASA's Earth science results, technologies, and data products to help resolve issues with regional economic and policy significance . . . supporting the U.S. Global Change Research Program" [11].

While V0 systems performed and interoperated efficiently and reliably at each DAAC, implementing a data management system of EOSDIS magnitude and generic capability was new to system developers, as well as system users. High costs, increasing system requirements, new technology cycles that were faster than development cycles, and trading innovation for predetermined requirements all pointed to the need for NASA to provide alternate opportunities to further advance the capabilities of data management systems. Internal to EOSDIS, "generation of standard products was moved, in most cases, to Science Investigator-led Processing Systems (SIPS)" [3], as exemplified by Cuddy et al. [12] (SIPS products were provided back to the respective DAAC), allowing DAACs to focus more on archiving, distribution, and data services, and PIs on the data (e.g., validation) itself. External to EOSDIS, the "federation of competitively selected Earth Science Information Partners (ESIPs)" [13], also known as the Federation of ESIPs, was created to foster community involvement in developing special research products, potential commercial products, and new data management technologies.

Also, in the late 1990s, NASA commissioned the "New Data and Information Systems and Services (NewDISS) Strategy Team to define the future direction, framework, and strategy of NASA's Earth Science Enterprise (ESE) data and information processing, near-term archiving, and distribution . . . . The main recommendations were to provide smaller, more heterogeneous components than were being developed with EOSDIS" [3]. NewDISS recommendations, the roots for the 2006–2007 evolution, were handed to the Strategic Evolution of ESE Data Systems (SEEDS) group, whose mission was to "Establish evolution strategy and coordinating activities to assure continued effectiveness of ESE data management systems and services" [14]. Four community-based working groups continue to derive information management infrastructure processes in developing standards, planning and reporting metrics, infusing new technologies, and reusing information management assets.

With the open availability of NASA Earth science data, it was recognized that advanced services to facilitate the access and use of the data were necessary for two driving purposes.

1) To expedite the data validation process, thus understanding the behavior of the retrieved remote sensing data and ensuing data processing better. With data being publicly distributed and used for research, decision support, and policy, it was essential that data be validated and understood in a much more timely fashion.

2) To gently introduce a new paradigm to the larger research community: the use of ordered and formatted satellite-generated digital data for the study of Earth science.

From the mid-1980s to the late 1990s, with each successive evolution study and activity, and technology advancement, information managers and technologists revealed increasingly more innovative information management ideas, technologies, and implementations to further evolve the use and access of data.

C. Age of Discovery

The ubiquitous availability of the World Wide Web (WWW), and the opportunities given to technologists and scientists to exploit this and other new information technology advancements, catapulted information management into an era of "discovering" new ways to manage data and information. Technology sharing bred implementation collaborations, resulting in more efficient science data sharing. Tools and services were invented to further facilitate data access, visualization, and analysis. More automated data management processes were implemented. For example, character user interfaces became graphical user interfaces, which later became Web-based user interfaces.

NASA's solicitation for Research, Education and Applications Solutions Network (*REASoN*) projects resulted in broad community collaborations that brought forth the most innovative ways to utilize NASA Earth science data. REASoN was followed by NASA's Advancing Collaborative Connections for Earth System Science (ACCESS) Program, further pushing the information technology community to integrate more efficient information management infrastructure and science data analysis tools. New ways of bringing data to the research community evolved, and partnerships to further understand the science data access needs developed between data managers and data users.

Currently, over 100 NASA-funded organizations advance data availability to information availability, thus contributing to Earth science information management evolution. Specifically, data centers became data and value-added information service centers. Services were implemented to enable data users to extract information out of the data. Other services included data subsetting, mining, visualization, preanalysis, fusion, and exploration, as well as rich meta-data (information about the data, including data history and versioning) and dynamic remote data access. With the free movement of data and information, scientists are able to create long-term Climate Data Records (CDRs) from information generated by several instruments.

Ackoff [15] identifies five categories to define the content of the human mind: 1) data; 2) information; 3) knowledge; 4) understanding; and 5) wisdom (see Fig. 1). Data centers have evolved far from not only providing data (having "no significance beyond its existence" [16]) to providing information ("data that have been given meaning by way of relational connection" [16]) to enable science.

Nowadays, the infusion of more sophisticated information management tools has evolved information management to the next phase: Earth science knowledge management ("knowledge is the appropriate collection of information such that its intent is to be useful" [15]). The rapidly increasing ability to analyze data from constellations of instruments, combining heterogeneous data sets, establishing processes for information management standards [17], enhancing data interoperability
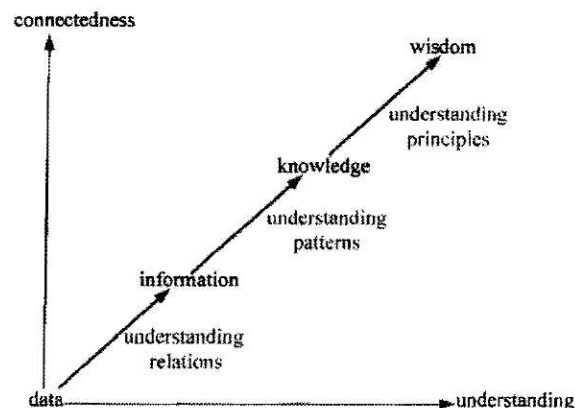

Fig. 1. Simplified content transition model [16].

[18], integrating affordable disk, and, perhaps, deploying true "knowledge building systems (KBSs)" [19], have all contributed to this further evolution.

III. EVOLUTION 2006–2007 SHAPING FORCES AND GOALS

The recent segment of the evolution continuum (2006–2007) represents a large step in the shift for NASA Earth science information management systems from providing information-bearing services to providing knowledge-bearing services. The, 2006–2007 evolution segment was driven by forces that made it one of the largest leaps in NASA information management evolution.

In early 2005, due to shrinking budgets for NASA information management systems unable to sustain ECS, aging ECS equipment, the need to be responsive to the implementation of new technologies, and the desire to make information management more distributed (and more closely tied to PI science teams), NASA embarked on an EOSDIS evolution study to develop a 2015 vision of information management systems. The effort, encompassing the results of previous committees, was comprised of "An EOSDIS Elements Evolution Study Team to provide an external viewpoint and offer guidance, and an EOSDIS Elements Evolution Technical Team to develop an approach and implementation plan that would begin to fulfill the objectives ... developed by the Study Team. The objectives that were part of the 2015 vision included: increasing end-to-end data system efficiency and autonomy while decreasing operations costs; increasing data interoperability and usability by the science research, application, and modeling communities; improving data access and processing; and ensuring continued safe stewardship .... This provided the tenets (and goals) (see Table I) under which the Technical Team conducted its analytical work" [18].

In what was known as Step 1, NASA approved the evolution of four information management systems as follows [20].

GES DISC: Consolidate GES DISC data holdings into one DISC-unique system. This featured transition of data sets generated by the Earth-observing instruments, namely, the Atmospheric Infrared Sounder (AIRS), the High Resolution Dynamics Limb Sounder (HIRDLS), the Ozone Monitoring Instrument (OMI), the Microwave Limb Sounder (MLS), and the Solar Radiation and Climate Experiment (SORCE), and

TABLE I
EOSDIS EVOLUTION 2015 VISION TENETS [5]

Vision Tenet	Vision 2015 Goals
Archive Management	• 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	• 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	• 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	• Mechanisms to collect and preserve the pedigree of derived data products are readily available.
Cost Control	• Data systems evolve into components that allow a fine-grained control over cost drivers.
User Community Support	• 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	• Access to all EOS data through services at least as rich as any contemporary science information system.

model data from the Global Monitoring and Assimilation Office (GMAO) to the Simple, Scalable, Script-based Science Product Archive (S4PA) data management system. Also, phase out ECS in 2007, and reduce archive volume, due to the transfer of the Moderate Resolution Imaging Spectroradiometer (MODIS) data (see below). (V0 data sets already reside in S4PA due to an earlier migration.)

Langley Research Center (LaRC) Atmospheric Sciences Data Center (ASDC) DAAC: Consolidate ASDC data holdings into one ASDC-unique system, namely, the Archive Next Generation (ANGe). This featured transitions of Cloud and Earth's Radiant Energy Budget (CERES) data from the heritage LaRC Tropical Rainfall Measuring Mission (TRMM) Information System (LaTIS) to the new ANGe archive.

PI-Led MODIS Adaptive Processing System (MODAPS): Transfer responsibility for MODIS processing, archiving, and distribution from GES DISC to MODAPS. This featured on-demand processing of precalibrated (level 1) data and closer involvement and control by the science community.

ECS: In parallel to the independent evolution at the three sites above, rearchitect ECS to simplify sustaining engineering and automate operations, to be deployed at the other DAACs where ECS is deployed. This featured a simplified software architecture.

All the evolved systems above will increase system automation and use online storage and commodity disks/platforms, thereby reducing operation, archiving, and system engineering costs, while providing quicker access to data. Thus, although the overall implementation at the first three sites (the data centers) did not drastically differ from each other, each system now would provide capabilities of specific interest to the community that it serves and would have different approaches to achieving this. ASDC's evolution approach was to "consolidate LaTIS and ECS, increase automation, leverage commodity hardware, build upon Open Source software, not impose changes to external data providers, and work with (local PIs) to leverage total resources" [21]. MODAPS's evolution approach was not to archive voluminous level 1 data, but to generate level 1 data on-demand, permanently archive Golden Month (the data set comprised of data processed for the same observation month/year,

but utilizing successive algorithm versions) and higher level data products, provide online search and order capabilities, and install direct access servers to data [22]. GES DISC's evolution approach was to consolidate data sets into discipline-specific archives, provide direct online access to data, remove Commercial-Off-The-Shelf (COTS) dependences and thus, integrations, reuse existing software, and build in flexibilities for future missions and community-driven enhancements.

With the evolution 2006–2007 underway, a set of information management system requirements to ensure that the evolved systems can perform functionally and efficiently to meet NASA's needs and commitments was defined [23].

Concurrent with the evolution study effort, many DAACs prepared for technical changes to their information management systems, to varying degrees, based on their knowledge of NASA evolution efforts. The GES DISC also recognized that EOSDIS at the GES DISC needed a technology refresh, and that the existing core of the EOSDIS, ECS, would not be affordable in its present architecture with drastically reduced budgets on the horizon. In response, the GES DISC evolution strategy was developed and, further, would be implemented with no additional funding by utilizing commodity hardware, exploring requirements retirement, and improving operational processes. The GES DISC strategies were consistent with those of the Evolution Study Team and were accepted as an implementation of the GES DISC approved Study Team recommendations, except for the need to process voluminous data on-demand (virtualize appropriate data), since the responsibility for these data (MODIS level 1) was moved to MODAPS. GES DISC evolution implementation will be discussed in Section IV.

IV. GES DISC EVOLUTION 2006–2007

As specified in the "Earth Observing System (EOS) Program Plan" [6], the EOSDIS is designed to provide a long-term data record for a broad range of environmental parameters. The mission requirements are to produce (or enable production of) standard science data products from EOS instruments (see [24]–[26] for instrument details), provide a distributed information framework supporting EOS investigators and other users,

TABLE II
ECS AND S4PA DATA MANAGEMENT COMPARISON

	ECS	S4PA
Data Management	Nearline, Tape Based	Online, Disk Based
Data Access	Orders	Download
Search Services	EDG, WHOM	WIST, Mirador, WHOM
Data Processing Platforms	SGI, Sun, Linux	Linux
Discipline Specific Archive	No	Yes
Operational Support	24 X 7	8 X 5

provide archiving and distribution services for data until they are transferred to long-term archives, provide seamless access to EOS data for discovery, search, and order, and interoperate with data archives of other agencies and countries.

The ECS, deployed at the GES DISC in 1999, is the component of the EOSDIS that provides the "core" common capabilities to meet EOSDIS requirements for spacecraft and instrument planning, scheduling, command, and control; and for science product generation, information management, data archival, and distribution (i.e., the data system). ECS ingested data from multiple data production systems, other EOS sources, and external data providers and managed the storage and access to these data. The majority of the data, growing to 2 PB at the GES DISC by 2005, was stored in tape-based archives and a smaller online "data pool," managed by a combination of COTS and custom software [5].

The GES DISC 2006–2007 evolution was focused on replacing ECS with the in-house developed disk-based S4PA data management system and migrating data to the disk archives. S4PA was previously used in the archival of pre-EOS era data sets. Transition from ECS to S4PA was handled incrementally. Each transition involved porting the archive to new Linux-based systems, performing an end-to-end interface test, running a performance test, running in dual operations with ECS, and transitioning following a successful Transition Readiness Review [27]. Transition was completed in December 2007, on schedule.

During the evolution, it was most important not to affect external interfaces, changes in the production system, access to data, or data stewardship. Thus, data provider interfaces for level 0 data and PI-produced products did not change, the Simple, Scalable, Script-based Science Processor for Measurements (S4PM) [28], a processing system similar to S4PA, continued to produce AIRS and value-added products, and support continued for the meta-data publication to the EOS ClearingHOuse (ECHO), enabling the science community to exchange data and information. As expected, however, many architectural changes to perform EOSDIS functions did occur. An ECS and S4PA architecture comparison is shown in Table II, and is described later in this paper. More specifically, the before and after evolution architectures are illustrated in Figs. 2 and 3.

In addition, the GES DISC evolution 2006–2007 enhanced community interactions from not only integrating data management tools in response to community needs, but also proactively seeking collaborations with community members seeking more sophisticated techniques to perform data analysis. For example, data visualization and access tools have been enhanced to accommodate aerosols and modeling scientists. Each evolved architectural component is described in the following.

Data Management: The rapidly dropping prices of disk storage allowed the GES DISC to deploy a purely online archive in the evolution system, where all the data are stored in an area accessible via an anonymous FTP (or HTTP for restricted data). Disk-based storage alleviates many of the operational costs from robotic tape archives, such as volume management and drive/tape/robot troubleshooting. In addition, it eliminates the need to map files to tape libraries and individual tapes. Also eliminated is the requirement to support asynchronous orders, where a user submits a batch request for data and receives an email when it is staged for pickup. Instead, the user can simply download the data directly from disk at any time. The elimination of asynchronous orders, in turn, obviates the need for order tracking. The removal of these complex components (volume management and order tracking) enables a much simpler architecture for managing the data, indeed, a storage architecture that is centered around the file system, rather than a relational database.

This simplified archive management system is S4PA, an offshoot of the earlier S4PM science processing system [28]. S4PA is a workflow-based system for managing science data ingestion, archiving, distribution, and meta-data publication. In addition, it incorporates continuous integrity checking, verifying file checksums as it cycles through all files in the archive every few days, another enhancement enabled by the online archive. Although the main archive is kept on disk, S4PA includes tape backups as well. The file system is structured into tape-sized partitions so that as each partition fills, a message is sent to the system administrators, who make a standard system backup tape for the partition.

Data Access: The online storage of data represents a significant change in the ways end users access data. The data are organized hierarchically, first by data set and then by data date. Thus, many users can simply navigate to the data of interest using a simple Web browser. Sophisticated users can (and do) also write scripts or applications to acquire data in bulk or on demand.

Just as importantly, the online access to the data enables the GES DISC to add a variety of synchronous data services that were possible only for a limited number of data sets before. These range from data-set-specific on-the-fly subsetting to standards-based services such as OPeNDAP [29], Open Geospatial Consortium (OGC) Web Map Service, OGC Web Coverage Service, and online analysis with Giovanni [30].

Search Services: Although many users could simply find and acquire the data by navigating the directory structure, a number of search services are provided to make this process easier. Prior to Evolution, EOSDIS included the EOS Data Gateway (EDG), which was able to search data at all the DAACs. This required the data centers to deploy a server-based search capability, typically backed by a database. However, with EOSDIS migrating from EDG to the Warehouse Inventory Search Tool (WIST) (which searches meta-data published in ECHO), S4PA has also migrated its EOSDIS search services support to WIST/ECHO.

In addition to EOSDIS-wide search, the GES DISC continues to offer local search interfaces for GES DISC holdings: a Web-based hierarchical navigation tool called the Web Hierarchical Ordering Mechanism (albeit with a nonanachronistic name) and a free-text search tool named Mirador [31]. Both tools allow a user to find the URLs for data matching specific
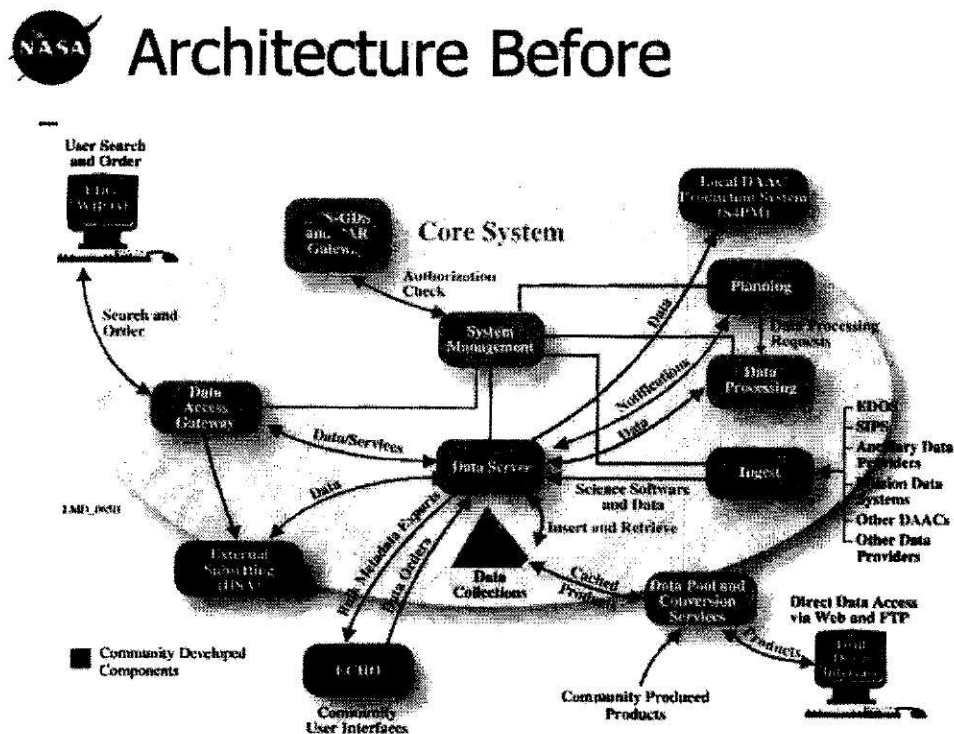
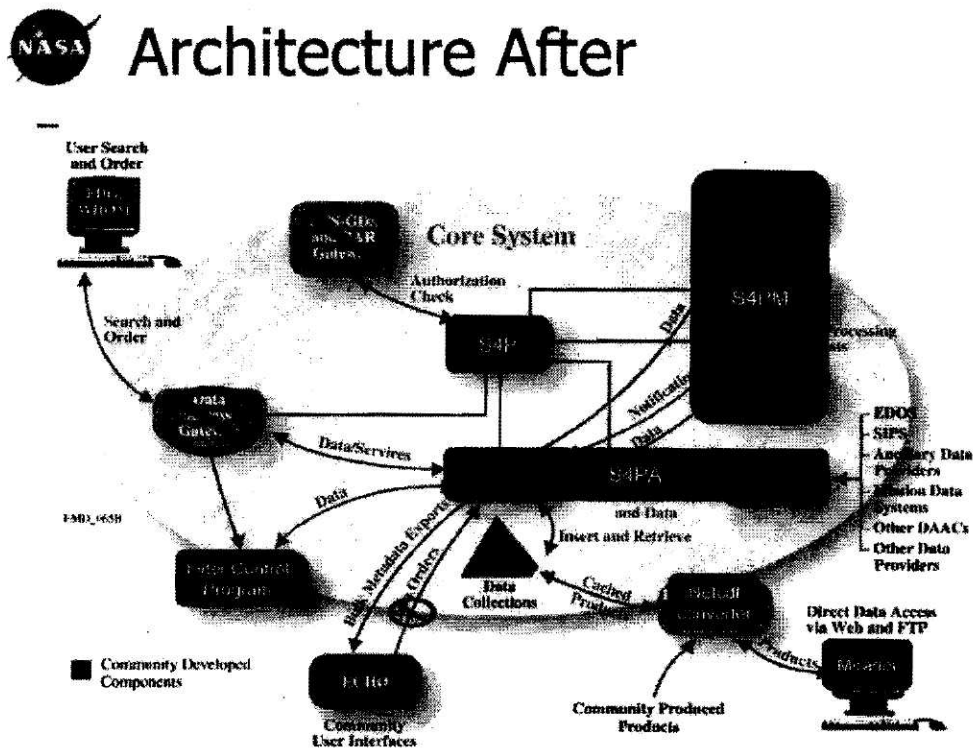Fig. 2. Architecture prior to GES DISC evolution.

Fig. 3. Architecture after GES DISC evolution (same functionality, simpler architecture).

data set, space, and time criteria, as well as access to on-the-fly services such as subsetting and format conversion. Furthermore, the Giovanni data visualization and analysis tool has become popular for data exploration and access. These services are hosted on machines separate from the main data servers (albeit connected via a high-speed network) so that processor-intensive queries and services do not interfere with the task of providing the actual data via FTP, and vice versa.

Data Processing: In addition to data management, the GES DISC also processes science data for some missions (Terra,

Aqua, and Aura), as well as valued-added products. S4PM continues to be the software system that runs this processing. The evolutionary change in this case is the migration of the processing from large Unix-based silicon graphics servers to multiple small commodity-class Linux servers. The low cost of individual commodity servers supports a more flexible provisioning strategy as well as more frequent technology refreshment. Processing is run on dedicated machines (or machines with negligible user access rates) so that unpredictable and uneven user loads do not interfere with standard processing.

Discipline-Specific Archive: Now that GES DISC resident data are archived on disks, data types can be broken out on different disk groupings and be sized accordingly. In fact, rather than support all missions and instruments in a single system, the S4PA architecture actually consists of several small standalone systems, each of them supporting a particular mission or set of measurements. This flexibility allows for discipline-specific services to be applied on certain data sets, which would be an unnecessary burden to support for other data sets. Thus, the one-size-fits-all paradigm is avoided. Furthermore, this enables tools to be developed that bring information together from multiple missions, thus contributing to enhancing knowledge and knowledge management. In addition, this "mini-data-center" approach has the effect of spreading out the user access load among multiple hosts, providing a certain level of load balancing and shielding the performance of individual systems against heavy loads resulting from high demand for any one data set. (In any case, individual systems have shown the ability to respond to demands of over 500 GB distribution per day with no noticeable effect on system operations.)

Operational Support: The S4P operator interface is markedly simpler than previous operator interfaces. There are fewer components to track, hardware will be relatively uniform, there are no tape silos to monitor, and there are no orders to manage. Operations staff need only be present 8×5, and are on call 24×7. In addition, preventive/corrective maintenance can be performed on each standalone system independent of the others. Thus, "system maintenance" need not affect all systems at once. Also, S4PM hardware is oversized to allow processing to catch up if downtime occurs during off hours.

Community Interaction: During 2006–2007, GES DISC interaction with the research community also evolved in response to the Evolution Study Team Vision Tenet, "User Community Support." The GES DISC, through conference and science team presentations, has proactively developed relationships with the research community seeking to initiate the use of information management tools to further advance Earth science knowledge. The flexibility and responsiveness of S4PA to implement new services has greatly facilitated support to the user community.

The benefits of GES DISC information management system evolution include reduction in operations costs due to elimination of multiple systems and reduction in sustaining engineering costs due to use of simpler scalable software, as well as reduction in the dependence on COTS products and their integration, increased system automation due to single system, simpler operational scenarios, and improved, more efficient, and less cumbersome data access. Perhaps the most significant benefit is the ability to layer information extraction services on top of the archived data.

V. CONCLUSION

According to GES DISC stakeholders (NASA Headquarters, the Earth Science Data and Information System (ESDIS) Project at NASA GSFC, and PIs), the GES DISC information management system evolution 2006–2007 successfully met the objectives approved by NASA. Matured and new technologies, and their dropping costs, have enabled the GES DISC to utilize the implementation paradigm successfully used for V0. This includes having experienced Earth science information management personnel develop an information management system that specializes in specific data sets, and thus, focuses on the needs of its community (with whom personnel are already familiar). The new EOSDIS at the GES DISC provides opportunities for improved system and operational efficiency, as well as enhanced data and information responsiveness and services. The ability of the new EOSDIS to integrate cutting-edge technologies and information management tools enables researchers to focus on science research, rather than on pre-science data preparation. This is exemplified by the A-Train Data Depot Project that provides services to bring together and coregister heterogeneous data sets, thus freeing researchers from individually repeating these tasks [32]. The feedback of more sophisticated prescience tools facilitating more advanced science research, which, in turn, would demand even more sophisticated tools, indicates that the knowledge management age has just begun. "Having a program requirement for continuous technology assessment establishes a culture of innovation ... Flexible management processes accommodating innovation, speed and efficiency are essential for increasing agility in development despite the higher perceived risks" [5].

Evolution continues at the GES DISC, after 2007, consistent with the Evolution Study Team Vision Tenet, for the continued purpose of enabling knowledge in Earth science. In addition to Mission Support, the GES DISC will focus on lines of business that include science data and information services, technologies that enable information management, multisensor data management, measurement-based information management systems, and community-requested data-brokering opportunities.

ACKNOWLEDGMENT

REFERENCES

[1] *NASA Earth Missions*, May 16, 2007. (last update). [Online]. Available: http://science.hq.nasa.gov/missions/earth.html
[2] *NASA Science Plan for NASA's Science Mission Directorate 2007–2016*, Jan. 2007.
[3] H. K. Ramapriyan, "NASA's Earth science data systems—Past, present and future," in *Proc. IGARSS*, Toulouse, France, Jul. 2003, pp. 637–639.
[4] National Academy of Science, National Research Council, *Earth Science and Applications From Space, National Imperatives for the Next Decade and Beyond*, 2007.
[5] M. Esfandiari, H. Ramapriyan, J. Behnke, and E. Sofinowski, "Evolving a ten year old data system," in *Proc. 2nd IEEE Int. Conf. SMC-IT*, 2006, pp. 243–250.
[6] *Earth Observing System (EOS) Program Plan*, Jul. 1, 2001. Doc No. 420-01-01, Rev A.

[7] P. Smith, L. Treinish, and L. Novak, *Pilot Climate Data System: A State-of-the-Art Capability in Scientific Data Management*, Oct. 1983. (abstract). [Online]. Available: http://ntrs.nasa.gov/search.jsp?R= 738679&id=3&qs=N%3D429485126

[8] J. Brown and C. Miller, *The NASA/JPL Pilot Ocean Data System*, Sep. 1984. (abstract). [Online]. Available: http://ieeexplore.ieee.org/xpls/ abs_all.jsp?arnumber=1152270

[9] R. Price and R. Ragan, *The Development of a Component of NASA's Pilot Land Data System to Support Hydrological Research*, 1986. [Online]. Available: http://www.cig.ensmp.fr/~iahs/redbooks/a160/iahs_ 160_0429.pdf

[10] NASA's Academy of Program and Project Leadership, *NASA EOSDIS Case Study & Lessons Learned*. Revised Jul. 2005.

[11] NASA Regional Earth Science Applications Centers, Oct. 22, 1998. Release 98-193. [Online]. Available: http://www.nasa.gov/home/hqnews/ 1998/98-193.txt

[12] D. Cuddy, M. Echeverri, P. Wagner, A. Hanzel, and R. Fuller, "EOS MLS science data processing system: A description of architecture and capabilities," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 5, pp. 1192–1198, May 2006.

[13] National Research Council, *A Review of the U.S. Global Change Research Program and NASA's Mission to Planet Earth/Earth Observing System*, 1995, Washington, DC: Nat. Acad. Press.

[14] S. Wharton et al., *Strategic Evolution of ESE Data Systems (SEEDS), Briefing to ESISS Subcommittee*, Jul. 10, 2003. (presentation). [Online]. Available: http://esdswg.gsfc.nasa.gov/Info.html

[15] R. L. Ackoff, "From data to wisdom," *J. Appl. Syst. Anal.*, vol. 16, pp. 3–9, 1989.

[16] G. Bellinger et al., *Data, Information, Knowledge, and Wisdom*, 2004. [Online]. Available: http://www.systems-thinking.org/dikw/dikw.htm

[17] "Development of the Earth science data systems standards process," *Strategy for Evolution of ESE Data Systems.* (SEEDS) Standards Process Group Page; last modified Dec. 16, 2007. [Online]. Available: http://esdswg.gsfc.nasa.gov/WG/spg.htm

[18] M. Esfandiari, H. Ramapriyan, J. Behnke, E. Sofinowski, "Evolution of the Earth Observing System (EOS) Data and Information System (EOSDIS)," in *Proc. IGARSS*, Denver, CO, Jul. 31–Aug. 4, 2006, pp. 309–312

[19] H. K. Ramapriyan et al., "Intelligent archive concepts for the future," in *Proc. ISPRS/Future Intell. Earth Observing Syst. Conf.*, Denver, CO, Nov. 2002.

[20] M. Esfandiari, *EOSDIS Evolution Plan*, Nov. 17, 2006. (presentation). [Online]. Available: http://edcdaac.usgs.gov/landdaac/presentations/ sap_03-06/EOSDIS_Evolution_Willems.pdf

[21] M. Ferebee, *EOSDIS Evolution at the NASA LaRC Atmospheric Science Data Center (ASDC)*, May 3, 2006. (presentation). [Online]. Available: http://science.larc.nasa.gov/ceres/STM/2006-05/Ferebee_ASDC0605.pdf

[22] E. Masuoka, "Evolution of data systems for processing, archiving and distributing global science products from the MODIS perspective," *Information Sciences & Technology Colloquium Series (abstract)*, Sep. 13, 2006. [Online]. Available: http://isandtcolloq.gsfc.nasa.gov/ fall2006/speaker/showcase.html

[23] *Requirements for Archiving, Distribution and User Services in EOS Data and Information System (EOSDIS)*, Apr. 2006, Greenbelt., MD: NASA/GSFC.

[24] K. J. Ranson, "NASA's EOS terra mission update," in *Proc. IGARSS*, 2003, vol. 5, pp. 3023–3024.

[25] C. Parkinson, "Aqua: An Earth-observing satellite mission to examine water and other climate variables," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 173–183, Feb. 2003.

[26] M. Schoeberl, A. Douglass, E. Hilsenrath, P. Bhartia, R. Beer, J. Waters, M. Gunson, L. Froidevaux, J. Gille, J. Barnett, P. Levelt, and P. DeCola, "Overview of the EOS aura mission," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 5, pp. 1066–1074, May 2006.

[27] M. Esfandiari, H. Ramapriyan, J. Behnke, and E. Sofinowski, "Earth Observing System (EOS) Data and Information System (EOSDIS)—Evolution update and future," in *Proc. IGARSS*, Barcelona, Spain, Jul. 23–27, 2007, pp. 4005–4008.

[28] C. S. Lynnes, "The simple, scalable, script-based science processor," in *Earth Science Satellite Remote Sensing*, vol. 2. Beijing, China: Tsinghua Univ. Press, 2006, pp. 146–161.

[29] P. Cornillon, J. Gallagher, and T. Sgouros, "OPeNDAP: Accessing data in a distributed, heterogeneous environment," *Data Sci. J.*, vol. 2, pp. 164–174, 2003. Available: http://www.jstage.jst.go.jp/ article/dsj/2/0/2_164/_article

[30] J. Acker and G. Leptoukh, "Online analysis enhances use of NASA Earth Science Data," *EOS Trans. Amer. Geophys. Union*, vol. 88, no. 2, p. 14, 2007.

[31] C. Lynnes, R. Strub, E. Seiler, T. Joshi, and P. MacHarrie, "Mirador: A simple fast search interface for global remote sensing data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 92–96, Jan. 2009.

[32] A. Savtchenko, R. Kummerer, P. Smith, A. Gopalan, S. Kempler, and G. Leptoukh, , "A-train data depot—Bringing atmospheric measurements together," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2788–2795, Oct. 2008.

Steven Kempler received the B.A. degree in geography from the State University of New York, Cortland, in 1976 and the M.A. degree in physical geography (atmospheric science) from The Ohio State University, Columbus, in 1979.

He is currently with the NASA Goddard Space Flight Center, Greenbelt, MD, serving as the Manager of the Goddard Earth Sciences Data and Information Services Center. His interests encompass the development of techniques, tools, and protocols for enhancing the use and usefulness of science data, and the development of new markets for Earth-observing remote sensing data.

Christopher Lynnes received the A.B. degree in earth sciences from Dartmouth College, Hanover, NH, in 1981 and the Ph.D. degree in seismology from the University of Michigan, Ann Arbor, in 1988.

He is currently with the NASA Goddard Space Flight Center, Greenbelt, MD, serving as the Chief Systems Engineer for the Goddard Earth Sciences Data and Information Services. His professional interests are search methods and scientific data mining.

Bruce Vollmer received the B.S. degree in earth science from Eastern Illinois University, Charleston, in 1981 and the M.S. degree in geography (climate science) from the University of Delaware, Newark, in 1985.

He is currently with the NASA Goddard Space Flight Center, Greenbelt, MD, serving as the Mission Support Manager of the Goddard Earth Sciences Data and Information Services Center. His interests include space-based remote sensing, science data processing systems, climate change research, and information extraction from remotely sensed data.

Gary Alcott received the B.S. degree in computer science from the Pennsylvania State University, University Park, in 1984.

He is currently with the NASA Goddard Space Flight Center, Greenbelt, MD, serving as the Operations Manager for the Goddard Earth Sciences Data and Information Services Center. His professional interests are in operations engineering in the areas of system monitoring and control.

Stephen Berrick received the B.S. degree in physics from Duquesne University, Pittsburgh, PA, in 1981 and the M.S. degree in astronomy from New Mexico University, Portales, in 1992.

He is currently with the NASA Goddard Space Flight Center, Greenbelt, MD, within the Earth Sciences Data and Information Services Center. He has ten years of experience in science data processing, remote sensing science data support, data management, data pattern recognition, and developing data processing systems and interfaces.