GODDARD SPACE FLIGHT CENTER

+ Visit NASA.gov

CISTO news
Computational and Information Sciences and Technology Office

+Home

CISTO News is the semi-annual newsletter of the Computational and Information Sciences and Technology Office of Goddard Space Flight Center's Sciences and Exploration Directorate.

CISTO News

- HOME
+ ARCHIVES

Editor
Mike Hollis

Associate Editor
Jarrett Cohen

Consultants
Lara Clemence
Jim Fischer
Nick Burke

Download a PDF version of the newsletter web pages:

+ CISTO News Summer 2007

FEATURE ARTICLES

High-End Computing at NASA: An Interview with Tsengdar Lee and Phil Webster
The leaders of the High-End Computing (HEC) Program and CISTO express their views on HEC's role within NASA and the impact of IT industry trends on computational science and engineering.
+ Read More
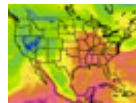
MERRA Project to Reconstruct Last 30 Years of Earth's Climate and Weather
The largest application ever hosted by the NASA Center for Computational Sciences will produce a comprehensive record of atmospheric conditions during the Earth-observing satellite era.
+ Read More

The Direct Readout Laboratory: Cost-Effective Support of NASA Satellite Missions Looking at the Earth
The Direct Readout Laboratory bridges the gap between older and newer technology satellites in a cost-effective manner.
+ Read More

NEWS UPDATES

Discover Cluster Expands with New Processors and Visualization Capabilities
With the addition of 1,024 processors, Discover will offer five times the computing power of the system it replaces.
+ Read More

PERSONNEL NEWS NOTES

Patrick Coronado—Recipient of Goddard's 2007 Excellence in Information Science and Technology Award
+ Read More

+ Privacy Policy and Important Notices
+ Sciences and Exploration Directorate
+ CISTO

Curator: Aimee Joshua
NASA Official: Phil Webster, CISTO Chief
Last Updated: Friday, 03-Aug-2007 12:21:03 EDT

GODDARD SPACE FLIGHT CENTER

+ Visit NASA.gov

CISTO news
Computational and Information Sciences and Technology Office

+Home

CISTO News

- HOME

+ ARCHIVES

Editor
Mike Hollis

Associate Editor
Jarrett Cohen

Consultants
Lara Clemence
Jim Fischer
Nick Burke

FEATURE ARTICLES

High-End Computing at NASA: An Interview with Tsengdar Lee and Phil Webster

By Jarrett Cohen and Mike Hollis

The *CISTO News* recently interviewed Tsengdar Lee and Phil Webster about the High-End Computing (HEC) Program's role within NASA and the impact of IT industry trends on computational science and engineering. Lee, HEC Program Manager, and Webster, CISTO Chief, are leaders in providing a service-oriented computing environment for NASA's Aeronautics Research, Exploration Systems, Science, and Space Operations missions. Both men began their careers as users of the high-end computers they now manage.

Lee received a PhD in Atmospheric Science from Colorado State University in 1992. He credits his training as a short-term weather modeler and later contributions to large information technology (IT) projects, including the National Weather Service's Advanced Weather Information Processing System, as guiding his present-day decisions. Lee joined NASA Headquarters in 2001 to manage the Scientific Computing and Global Modeling and Analysis Programs for the Earth Science Enterprise. Since July 2004, he has served within the Science Mission Directorate as HEC Program Manager. HEC encompasses the NASA Advanced Supercomputing (NAS) Division at Ames Research Center and the NASA Center for Computational Sciences (NCCS) at Goddard Space Flight Center.

After earning a PhD in Mechanical Engineering from Penn State in 1986, Webster specialized in computational aerodynamics for the Air Force Research Laboratories. He then spent 5 years with the Department of Defense's High Performance Computing Modernization Program (HPCMOD), where his varied experiences included managing large acquisitions, directing strategic planning, and providing oversight to HPCMOD computing centers. Webster joined the NCCS in 2000 and became CISTO's Lead for High-Performance Computing in 2005. In August 2006, the Sciences and Exploration Directorate appointed Webster as CISTO Chief. CISTO spearheads Directorate activities in scientific computing, networks and IT security, and information science and technology research.
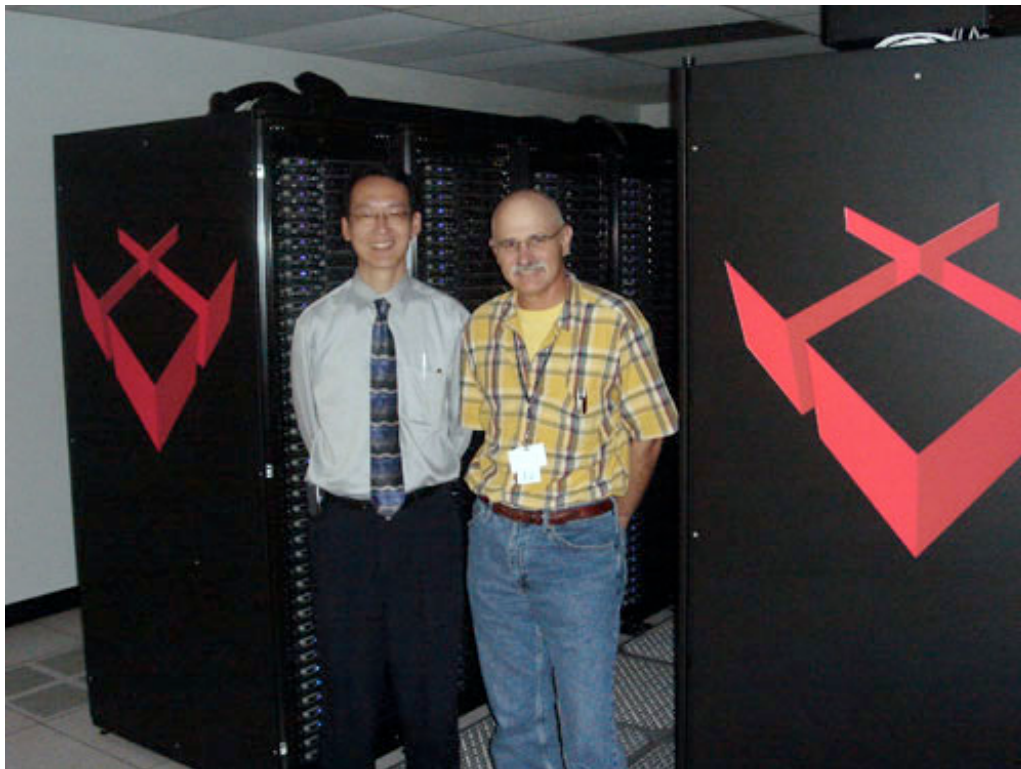
Tsengdar Lee and Phil Webster lead High-End Computing Program efforts at NASA Headquarters and Goddard Space Flight Center, respectively. They are standing near "Discover," the newest computing system at Goddard's National Center for Computational Sciences.

What follows are excerpts from Lee and Webster's wide-ranging responses to the *CISTO News* (CN) interviewers' questions.

CN: How is the HEC Program viewed among your Headquarters colleagues?

Lee: From the Earth science side, it is paramount. We cannot do that kind of research without a computational capability. Moreover, we are seeing increasing utilization from the space science side of the Science mission. In addition, our Aeronautics mission is focusing more on fundamental aeronautics, and much of that work is computationally intensive. Similarly, when they design a vehicle, Exploration needs to use very complex models to simulate vehicles through different flow regimes. So, HEC is very, very important to all NASA missions.

CN: Are there any missions that need high-end computing in space? The James Webb Space Telescope, which wants to unfold a large aperture and align it properly, comes to mind.

Lee: Right now, there has been a lot of discussion of high-end computing in space. That is related to data assimilation, trying to extract information from the observations and send the minimum amount of bits down to the Earth. Since you mentioned unfolding an aperture, in fact we are looking into some engineering problems. For example, when SOFIA's [Stratospheric Observatory for Infrared Astronomy] aperture creates a cavity, what kind of aerodynamic impact is there on the spacecraft? Also, there are interesting ideas to use SOFIA to observe near-Earth objects. We need to block or shield the starlight, and there has been discussion of using supercomputing to simulate that kind of shield.

CN: How robust is the Headquarters funding climate with respect to HEC?

Lee: Actually, it is quite stable. We had an increase in funding last year, and we are trying to maintain that. With Headquarters funding, NASA programs, overall, are under a lot of stress. High-End Computing is one area that we are trying to protect. That is a very significant commitment to maintaining a robust HEC capability.

CN: What recent or current HEC applications have had a major impact on NASA missions? In the near-term, what high-impact applications do you envision?

Webster: I see a tremendous amount of work that has come out of Jim Hansen's Goddard Institute for Space Studies and their IPCC [Intergovernmental Panel on Climate Change] simulations. A lot of interesting and ground-breaking work has come out of Joan Centrella's group with the black hole research they have been doing. Climate model development in the GMAO [Global Modeling and Assimilation Office] is very interesting and will have a large impact.

Lee: Exploration and Space Operations do a lot of engineering and CFD [computational fluid dynamics] research. Also, they are looking at how to design heat shields and use different materials that will sustain flight through the very difficult environment of space. These materials go through very different flow regimes. Furthermore, when you put people in the Crew Exploration Vehicle, how do you protect them from radiation? Many computing cycles are necessary to simulate space radiation and its interaction with spacecraft materials.

CN: What do you anticipate or hope will be the major scientific or engineering advances that result from NASA HEC?

Webster: Thirty or forty years ago, people had two tools to do research and engineering. One was theory, and the other was experimental observations. And now, with the advent of HEC, we have the ability to do simulations with incredibly high temporal and spatial resolution. I believe that we will be able to fully simulate what is happening with the climate. The exact causes of climate change and what, if any, mitigating actions should be taken are different than NASA's scientific mission of being able to predict. That becomes a political or a policy issue for other people within the federal government. NASA is an incredibly great place to work, and it has accomplished incredible things in the past. NASA has a very ambitious agenda of exploration and research. I believe that high-end computing is going to be a critical component of the entire Exploration mission. I don't see how we can develop vehicles, missions, and robots to go to the far corners of the universe without simulation on high-end computers.

Lee: With more and more observations at higher resolution, I see an opportunity coming up when we use high-resolution simulations and data assimilation to validate or to prove theories. I see that new, significant use of high-end computing in the future.

Webster: There are many fundamental questions of basic physics that, at some level, have to be modeled. If you look at Joan Centrella's research, the basic interactions of monstrous black holes, there are fundamental things in there that can only be fully understood with using high-end computing.

Lee: With Joan Centrella's work, a major question being addressed is if we can simulate the observations and try to predict what we can observe once we put LISA [the Laser Interferometer Space Antenna] in space. Similar work is starting to happen in Earth science with OSSE, the Observation System Simulation Experiment. You simulate your observation in a virtual world and use that as a design and validation tool.

CN: What is your vision for a future NCCS and HEC Program?

Webster: My vision for the future of the NCCS is what I call a data-centric computing center. High-performance computing centers in the past have been primarily processor-centric. Within the NASA community, new knowledge comes from the data. So, I'm trying to build the NCCS around data and data services that allow researchers to interact with and interrogate very large data sets, whether generated numerically or experimentally derived.

Lee: Data are really what we are about at NASA. We put up a lot of space assets, we collect data, and we generate a lot of data, so we have to focus on data. In addition, if we look at history, supercomputing centers did a lot of research building new supercomputers. Now, high-end computing is very quickly becoming a commodity. In the future, we need to develop expertise in acquisition, service, management, and exploring new offerings from the market instead of building our own systems.

CN: As PC clusters become more powerful and affordable, how will this impact traditional HEC as viewed by NASA? That is, do you anticipate a continuing need for centralized nodes at Ames and Goddard?

Webster: Historically, a tremendous amount of work was done on workstations, but it was always limited to the size of the workstation. Now, I believe these PC clusters have nominally replaced the workstation, but they still will not approach the size of problems that will be supported on supercomputers. So, supercomputers or high-end computers will always be there and will always be beyond what a single cluster, user, or project can do. Our scientists will demand a computer at that level for the advancement of science.

So, what is the impact on NASA? These PC clusters are still being built out of commodity processors, but they are growing. They are incredibly power-hungry, and they generate a lot of heat, which means that you can't just put them anywhere anymore because they put a tremendous burden on a building's infrastructure. So, if nothing else, a place like the NCCS or NAS at Ames has the infrastructure to support these types of computers. As a matter of fact, our limiting issue is the amount of cooling available for supercomputers that are projected to be deployed in the country in the coming years. Bottom line, we are driven to the point that we must use the components the industry provides for us. However, we will always be configuring these machines well beyond the design envelope that they were designed for, and hence the challenge of supercomputing.

Lee: I don't really see PC clusters, per se, becoming widely used. But the type of systems Phil just described, you take the most capable server from the market and try to configure it beyond what the commodity market will do. You take 16- or 32-way servers and put them together and make a bigger cluster. That is the way we push the limit. The time has passed for us to design our own computer from the ground up. We cannot afford to do that anymore.

Also, if you look at technology advancement, it is following Moore's Law [the number of transistors in an integrated circuit doubles every 2 years, with component cost staying the same] and exponentially increasing. If you think of human intelligence, we are not advancing or growing at that kind of speed. It is challenging for scientists and engineers to design applications to take full advantage of the most sophisticated systems. For NASA, that challenge hinges on whether or not you can analyze the data, make sense of the data, and also find the observations to match up with what you can simulate. A computer is a very good tool to generate data by itself, but you have to constrain that data with your observations. If you have the most sophisticated simulation but don't have data to validate or constrain it, what is good about it?

CN: What value beyond computing cycles does HEC provide to its users?

Webster: If you look at how these computers are being built, we are at a paradigm shift, moving from ever-increasingly powerful single processors to multiple processors all running at the speed of the old ones. There are a lot of issues about how users program these multi-core processors and, as Tsengdar described, how they deal with and evaluate the data. So, HEC needs to be able to supply knowledge to users about how to program these machines. We also have to provide a larger analysis environment and a visualization environment, things we have not traditionally done. As the problems get bigger and the data sets get so large, we have to present different tools to the user to interrogate these data.

CN: Some say that the network is the computer. How important is this issue to NASA HEC?

Lee: It is absolutely critical. If we look at the direction of HEC in the use of commodity systems and the building up of the clusters, the network is used for interconnect inside the cluster. The network also connects the compute systems to analysis nodes and mass storage available locally—as well as remotely. The network will be especially critical in the future, when not only will the compute nodes be distributed but also the data. You will have data sets in different locations, and you will have to bring those data to the analysis nodes.

Webster: I agree philosophically with the statement, but I don't believe the statement is correct that the network is the computer. Factually, you would have to say that the network is the backplane of the computer. If you compare your standard workstation, which has disk, screen, mouse, and processor, it also has to have a backplane that connects them. So, all the things Tsengdar just talked about, with a high-speed network as the backplane, now you can have the components of the computer geographically dispersed.

CN: What benefits to HEC users do you foresee from the transition to the NASA Integrated Services Network (NISN) away from the current NASA Research and Engineering Network (NREN)?

Lee: It is an economy-of-scale kind of argument. The Agency has a service-provider in NISN, so why don't we leverage it instead of developing our own dedicated resource? In terms of the service aspects, NREN and NISN are not too different in their offerings. Of course, there are different technologies, and we can argue whether this technology is better than that one. But, strategically, NASA is in a more budget-constrained environment. We have to leverage other organizations' offerings, just as we share our own resources with other organizations.

CN: Are you satisfied with the organizational structure of HEC at both Ames and Goddard? Do you anticipate the need for any reorganization?

Lee: That is really up to the Centers.

Webster: Certainly within Goddard, I believe that the NCCS will probably need to be more aligned in the future with the Applied Engineering and Technology Directorate, Code 500. We have a very robust engineering activity here. I suspect that there will be reorganizations of our computing center, both for alignment with the missions as well as the changing focus of the Information Technology and Communications Directorate, Code 700.

CN: Is the Engineering Directorate knocking on your door for compute services?

Webster: Not yet, but they have only recently started knocking on Tsengdar's door as Exploration has fulfilled itself. They now have a strategic plan, and they are starting to develop a research agenda and programs. I expect that, as Goddard becomes more competitive for a lot of these future missions, high-end computing will become necessary. If I just look at the pioneering work that Rick Lyon is doing in computational optics, the next step is modeling the different parts of a spacecraft.

http://www.hec.nasa.gov

+Home

CISTO News

- HOME

+ ARCHIVES

Editor
Mike Hollis

Associate Editor
Jarrett Cohen

Consultants
Lara Clemence
Jim Fischer
Nick Burke

FEATURE ARTICLES

MERRA Project to Reconstruct Last 30 Years of Earth's Climate and Weather

By Jarrett Cohen

MERRA, the largest application ever hosted by the NASA Center for Computational Sciences (NCCS), will soon be running on an augmented "Discover" Linux cluster. The Modern Era Retrospective-analysis for Research and Applications, as MERRA is formally known, will consume 544 processors of the NCCS's Linux Networx Custom Supersystem for 18 to 24 months.

As its name implies, a retrospective analysis, or reanalysis, starts in the past and marches forward in time to computationally reconstruct atmospheric conditions. Reanalyses are commonly undertaken by climate and weather centers, which use the same data assimilation systems as they apply to their forward-looking forecasts.

MERRA is an endeavor of the Global Modeling and Assimilation Office (GMAO) at Goddard Space Flight Center. One of the GMAO's primary functions is to support NASA Earth Observing System (EOS) satellite instrument teams and field experiments with assimilated data products in near real-time. The GMAO also uses models and assimilation systems to document and understand climate variability and predictability.

Traditionally, weather centers have an archive of analyses built up from their real-time systems, which are continually changing as upgrades occur. In contrast, the GMAO "can go back with a reanalysis and have a consistent system over time for doing the processing," said Michael Bosilovich, GMAO meteorologist and MERRA co-principal investigator with fellow GMAO meteorologist Siegfried Schubert.

With MERRA, GMAO scientists particularly hope to gain new insights into Earth's water cycle and understand how it changes with underlying climate variability. Research community access to the resulting data will expand the prospects for advances.

Data assimilation systems pair a computer model with an analysis that ingests, calibrates, and quality controls observations and then feeds them to the model. "Observational products are discontinuous in space and time," Bosilovich said. "Where we don't have data, the model fills in the gaps." Together, he explained, model and analysis provide globally gridded meteorological information that is uniform and reliable.

The GMAO is currently doing production-scale testing of the next version of its Goddard Earth Observing System-5 Data Assimilation System (GEOS-5 DAS), planned for MERRA. GEOS-5 DAS incorporates the GEOS-5 atmospheric general circulation model and the Gridpoint Statistical Interpolation (GSI) analysis. GSI was developed by the National Weather Service's National Centers for Environmental Prediction (NCEP) with GMAO contributions.
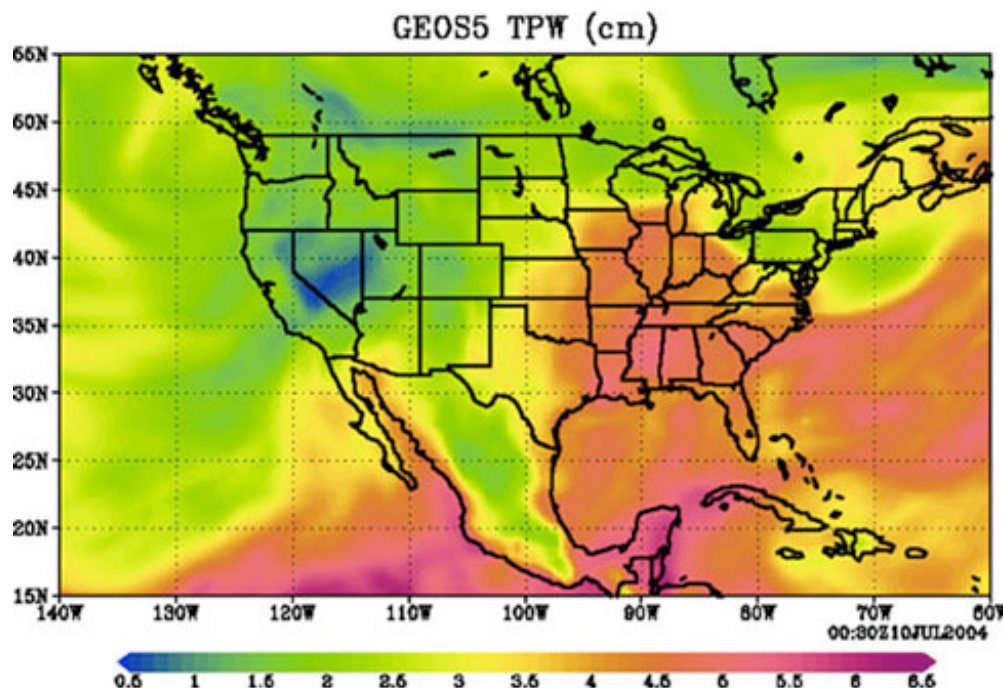
This visualization shows Total Precipitable Water (TPW) in centimeters over the United States on July 10, 2004. The data come from an experimental reanalysis of Summer 2004 running the GEOS-5 Data Assimilation System (DAS) at the 1/2-degree resolution being used for MERRA.

Reanalysis at a new scale

By using the GEOS-5 DAS, MERRA is leveraging the GMAO's most capable analysis tool to date, as well as the researchers developing it.

MERRA will produce a comprehensive record of Earth's climate and weather from 1979, the beginning of the Earth-observing satellite era, up to the day it finishes sometime in 2009. MERRA will cover twice as much time as the first GEOS reanalysis, which the predecessor Data Assimilation Office conducted with the GEOS-1 assimilation system back in 1993.

In addition, MERRA will have a horizontal resolution of 1/2-degree—roughly one model grid point every 55 kilometers—and 72 vertical levels. That compares to the 1-degree resolution used by the most recent large-scale reanalysis, the Japanese 25-year Reanalysis Project that ended in March 2006. Bosilovich said that finer resolution better locates observational inputs geographically, improving precipitation and other details in the analysis.

The NCCS computational offerings drove the GMAO's decision to alter their original plans for 1-degree resolution and 36 levels. "Since the computational capability is there, we are running MERRA at this higher resolution," said Gi-Kong Kim, GMAO production group lead. "Five years ago, such high spatial resolution was inconceivable."

MERRA also will have fine temporal resolution, with much of the diagnostic data produced every hour. Other diagnostics will be released at 3-hour intervals, which is typical for recent reanalyses. Among MERRA's more than 300 diagnostic variables are temperature, moisture, wind, surface pressure, and— unique with reanalyses—fields for the chemistry transport community. Hourly diagnostics will resolve the diurnal (daily) cycle of minimum and maximum values more precisely than existing 3- and 6-hourly reanalysis data products. They also will allow close study of individual weather events that can pop up within a few hours.

"Since it appears early in production, I am especially interested to see the Presidents' Day Storm of February 1979," Bosilovich said. This storm spread record-breaking snow amounts from the Ohio Valley to the Mid-Atlantic, with snowfall rates greater than 4 inches per hour around Washington, DC. At the time, weather models failed to predict the storm's intensity. Bosilovich also mentioned the 1988 Central U.S. droughts and floods, the 1993 Midwest floods, and several major El Niños as events of interest.

Besides increased spatial and temporal resolution, the variety and number of observations that MERRA must assimilate is much greater than in the previous reanalysis. Observation sources include satellites, ground stations, weather balloons, ships, and aircraft. "GSI can handle a lot more observations per day than in the past," Bosilovich said. He said this capability is especially important for assimilating 151 channels from the Atmospheric Infrared Sounder (AIRS) instrument on board the EOS Aqua satellite. AIRS represents 40 percent of the observations to be assimilated since Aqua's 2002 launch. For the Aqua

period, the GSI will read in 8 million observations per day for potential inclusion in the assimilation. Through quality control and other thinning, 5 million of those finally get assimilated into GEOS-5.
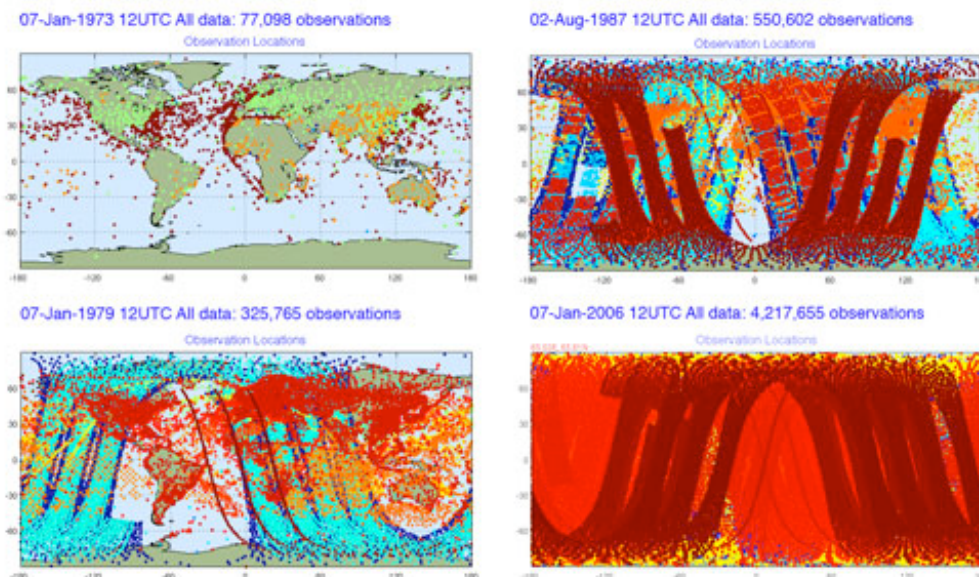
The number of Earth observations to be assimilated has increased dramatically over the last few decades. The panels demonstrate the evolution of observing systems from 1973 (pre-satellite) to 1979 (TIROS Operational Vertical Sounder [TOVS]) to 1987 (add Special Sensor Microwave Imager [SSMI] and several TOVS) to 2006 (add Atmospheric Infrared Sounder [AIRS] and several each of TOVS and SSMI). The headers list the number of observation points for an example 6-hour period during each year.

Production strategies

These new scale factors translate into substantial computing and storage requirements during MERRA's production phase. For example, doubling the spatial resolution increases processing needs eight-fold. Add the factors together, and requirements grow exponentially.

As the MERRA design expanded, it became clear that no current supercomputer could analyze all 30 years in sequence and finish within the 18-to-24-month window. To meet that timetable, MERRA must analyze about 30 "data days" per wall-clock day. Consequently, the GMAO divided the reanalysis into three concurrent streams of 10 years each. By running on 128 Discover processors, each stream can analyze 10 to 11 data days per day. An additional 128 processors will run supporting experiments.

All MERRA output will reside in the NCCS mass storage system. With an estimate of 15 gigabytes of output per data day, expected total data volume is 165 terabytes.

"The NCCS has done a very good job in having its production systems ready," Kim said. "The GMAO and NCCS have been working together very closely to keep the computing requirements updated."

The original plan was to run MERRA on "Explore," a 1,152-processor SGI Altix 3700 system. After test runs showed GEOS-5 performing faster on Discover, the GMAO decided to take advantage of the cluster's expansion to 2,560 processors this summer (see "Discover Cluster Expands with New Processors and Visualization Capabilities" in this issue). They also wanted to ensure completing the reanalysis on one supercomputer. "Given the length of the run, Explore's service contract might run out," Bosilovich said. Citing a large weather center that had to stop a reanalysis because the computer went away, he stressed the need for "continuity in computing systems and a stable platform."

The Discover expansion includes not only 1,024 new processors but also an additional 120 terabytes of online disk, which will better support MERRA production assessment and post-processing. Specialized Visualization Nodes enable close monitoring of the output streams. To speed data transfer between the mass storage system and Discover, the NCCS technical staff increased the network pipes to 10 gigabits per second. Staff also reconfigured the mass storage system to accommodate MERRA input/output, which is particularly write-intensive, explained Harper Pryor, CISTO Programs Development Manager (SAIC).

For the GMAO, Kim said one of the most challenging aspects about preparing for MERRA production has been collecting, identifying, and preparing all the input observations. Over several months, the GMAO made a "data sweeper" run of the reanalysis at 2-degree resolution using Explore. "The purpose was not so much to validate the scientific product but to ensure that the input observations would not give us surprises while we do production," Kim said. "If we encountered GEOS-5 system crashes due to problems related to input observations, we investigated and then designed solutions for production." To finish as quickly as possible, the sweeper also used multiple streams.

The final preparatory step is validation studies, which began running on Discover during July. "We are making sure the science has been applied properly," Bosilovich said. With the water cycle receiving prime attention, one of the comparative data sets is from the Goddard Laboratory for Atmospheres' Global Precipitation Climatology Project. Additional validation data come from remotely sensed data products (including MODIS, the Moderate Resolution Imaging Spectrometer on Terra and Aqua), other reanalyses, and ground stations. Such variety will enable verifying correct large-scale and regional circulation in the atmosphere.

Review of validation results is the responsibility of an External Users Group of scientists from universities, other climate and weather centers, and NASA field centers. "Rather than determining on our own when we are or are not ready to start production, we assembled an External Users Group to evaluate community needs," Bosilovich said.
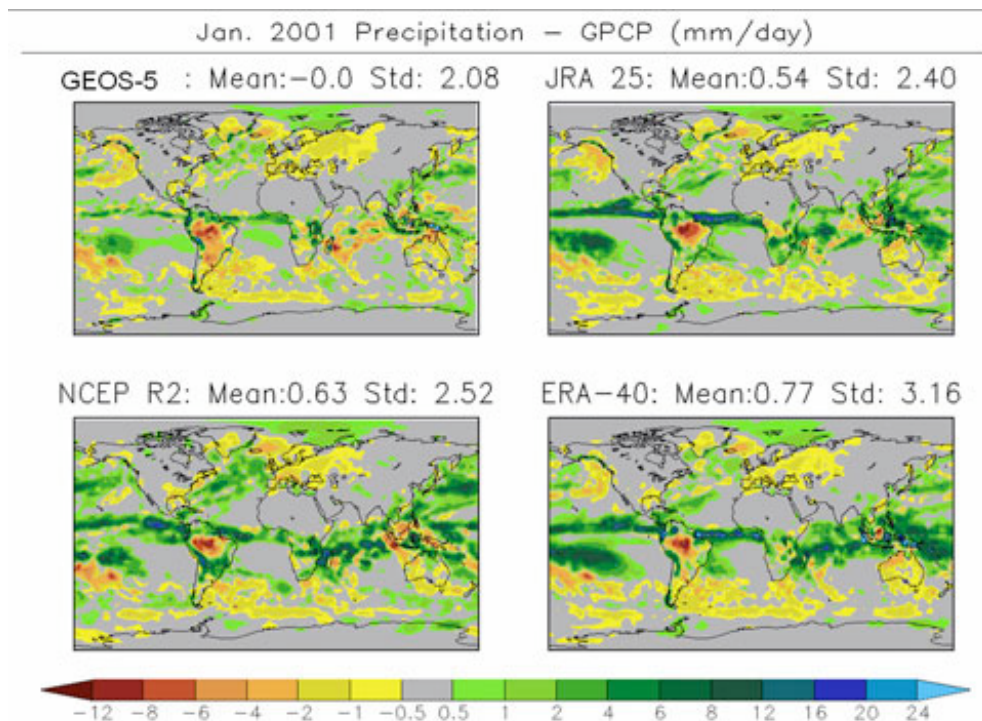
MERRA validation studies compare merged satellite and in situ observations from the Global Precipitation Climatology Project (GPCP) to reanalysis data. The panels show the differences in monthly mean precipitation (in millimeters per day) between GPCP and data from the GEOS-5 DAS and three completed reanalyses: the National Centers for Environmental Prediction Reanalysis-2 (NCEP R2), the European Centre for Medium-Range Weather Forecasts 40 Year Reanalysis (ERA-40), and the Japanese 25-year Reanalysis Project (JRA 25). GEOS-5 has a mean of zero and the smallest standard deviation, indicating the best agreement with GPCP. Of particular note is GEOS-5's improved tropical precipitation bias compared to the reanalyses.

Sharing the results

The broader Earth science community will have online access to a subset of MERRA data through the Goddard Earth Sciences Data and Information Services Center (GES DISC), formerly the Goddard Distributed Active Archive Center (DAAC). The GMAO estimates the available data will equal 55 terabytes, including the chemistry transport fields

The NCCS will funnel the data to GES DISC before archiving to tape, so scientists will have access to provisional data while MERRA is running. In collaboration with the GMAO, GES DISC is building a disk-based data storage and distribution system with a Web user interface. Kim said that planned capabilities include search using a variety of parameters, on-the-fly subsetting, and visualization to aid data search and comparison.

MERRA has a range of potential applications. For one, a researcher may pick case studies from the historical record. Because MERRA spans several generations of Earth-observing platforms, the GMAO will be able to look at climate variability and predictability from a uniform perspective. This research is part of their contribution to the U.S. Climate Change Science Program. Modeling groups could also use the data to drive mesoscale/regional models or global chemistry models. "Transferring a scientific product to the community has been our goal since writing the proposal," Bosilovich said. "That will be exciting to see in action."

MERRA funding comes from NASA's Research, Education, Applications Solutions Network (REASoN) and

Modeling, Analysis, and Prediction (MAP) programs.

http://gmao.gsfc.nasa.gov/research/merra
http://disc.gsfc.nasa.gov

+ Privacy Policy and Important Notices
+ Sciences and Exploration Directorate
+ CISTO

Curator: Aimee Joshua
NASA Official: Phil Webster, CISTO Chief
Last Updated: Friday, 03-Aug-2007 11:46:18 EDT

NASA | GODDARD SPACE FLIGHT CENTER

CISTO news
Computational and Information Sciences and Technology Office

+Home

CISTO News

- HOME

+ ARCHIVES

Editor
Mike Hollis

Associate Editor
Jarrett Cohen

Consultants
Lara Clemence
Jim Fischer
Nick Burke

FEATURE ARTICLES

The Direct Readout Laboratory: Cost-Effective Support of NASA Satellite Missions Looking at the Earth

By Mike Hollis

There are literally thousands of satellites orbiting the Earth at the present time. About 20 are active NASA Earth Observing System (EOS) satellites dedicated to measuring the planet's atmosphere, oceans, land, ice, and life. These measurements are more accurate, more frequent, and of higher spatial and spectral resolution than ever before. The EOS satellites capture massive amounts of data taken across a vast spectrum of wavelengths and simultaneously transmit these data to ground stations by omni-directional radio broadcasts at L (~1.5 GHz), S (~2 GHz), and X (~8 GHz) band frequencies. There are approximately 150 direct readout ground stations that intercept live X-band broadcasts in the satellites' lines of sight at the highest data rate of ~15 Mbps. These stations enable NASA and university scientists, commercial enterprises, governmental agencies, and other entities to use EOS data for their own interests, which include climate change studies, education, hazards management, agriculture, health, transportation, and geology, to name just a few.

While the Direct Readout Laboratory (DRL) here at Goddard Space Flight Center is just one of the ~150 ground stations with similar technological capabilities, the DRL serves a unique purpose with respect to all other stations. Under the direction of Patrick Coronado (see Figure 1), the DRL bridges the ground station readout gap between older technology satellites and upcoming new technology satellites. For example, as in-orbit satellite technologies change of necessity, it is the primary role of the DRL to make that transition cost effective for other direct readout ground stations. The DRL has formal representation on any new EOS satellite mission being planned in order to provide mitigating input to new spacecraft and instrument designs that could cause severe cost and/or technical problems for existing ground stations. As Coronado puts it, "Sometimes a 'better' in-orbit design feature suggested by instrument and spacecraft engineers can be the enemy of a 'good' improvement that doesn't create major cost problems for all the ground systems."

Figure 1: Patrick Coronado, Manager of the Direct Readout Laboratory (DRL), shown with the 2.4-meter diameter X-band development antenna on a mobile platform.

Occasionally, the DRL effects the transition from one generation of satellites to another with hardware improvements, but more often new software designs by the DRL staff are needed to upgrade the other ground systems. Further, at approximately 18-month intervals, the DRL holds meetings here at Goddard for representatives of the other direct readout ground stations to communicate information about changes, technologies, policies, and problems that affect the loose confederation of stations. Since the DRL is on the leading edge of direct broadcast and direct readout issues and helps make direct readout ground systems more cost effective, NASA Headquarters funds this effort along with the current EOS Program (in particular, the Terra and Aqua satellites) and its successor, the National Polar-Orbiting Observational Earth Satellite System (NPOESS) Preparatory Project (NPP). The NPP is the vehicle for providing technology infusion into the NPOESS program, which is a tri-agency effort of NASA, the National Oceanic and Atmospheric Administration (NOAA), and the Department of Defense (DOD).

The Terra polar-orbiting satellite, launched in 1999, is continually mapping the Earth's land surface and measuring heat, light, and the atmosphere to record environmental changes and human impact on climate. The Aqua polar-orbiting satellite, launched in 2002, continuously maps the Earth's water system to study the effects of heavy weather, hurricanes, tornadoes, snowstorms, etc., providing insight into the water cycle of evaporation and precipitation. NPOESS merges the interests of NASA, NOAA, and DOD into a constellation of three polar-orbiting satellites. Each satellite is designed for a 7-year lifetime and will host 10 to 12 sensor payloads that will provide a wide range of weather and environmental data. The first NPOESS launch is scheduled for 2012.

The DRL can capture data from all NASA EOS satellites and many non-NASA satellites. However, the DRL has NASA mission responsibilities to capture and send level-zero data from the Moderate Resolution Imaging Spectro-radiometer (MODIS) instruments aboard Terra and Aqua to the Goddard Earth Sciences Data and Information Services Center (GES DISC), formerly known as the Goddard Distributed Active Archive Center (DAAC). The importance of level-zero products is that these are raw data directly output from instruments at full resolution, and the raw data provide the basis from which all other data products are produced. For example, MODIS data provides a means for quantifying land surface characteristics such as land cover type, snow cover, surface temperature, leaf area index, and fire occurrences.

While these EOS satellites have their own independent ground system in the EOS Data and Operations System (EDOS), the DRL effort is important to the NASA mission in two respects: First, the DRL data set provides a check and data backup for any problems incurred with EDOS data storage or transmission. For example, the normal route along the EDOS ground system could be corrupted due to some hardware fault, and the DRL route provides an alternative means to obtain good data. Here it must be emphasized that the DRL only captures broadcasts from Terra and Aqua that are directly in the line of sight of its production antenna at Goddard (see Figure 2). The extreme range of this local line of sight extends approximately 2,000 km east and west and 3,000 km north and south of the antenna. Of course, instruments from these satellites are taking data constantly in a downward-looking mode, and data can be missed outside of the DRL local area, i.e., the East Coast vicinity of the United States. Second, the DRL provides a service to

the EOS mission in these cases of missing data by locating and pulling good data from any of the 120 other direct readout ground stations and supplying it to the GES DISC.
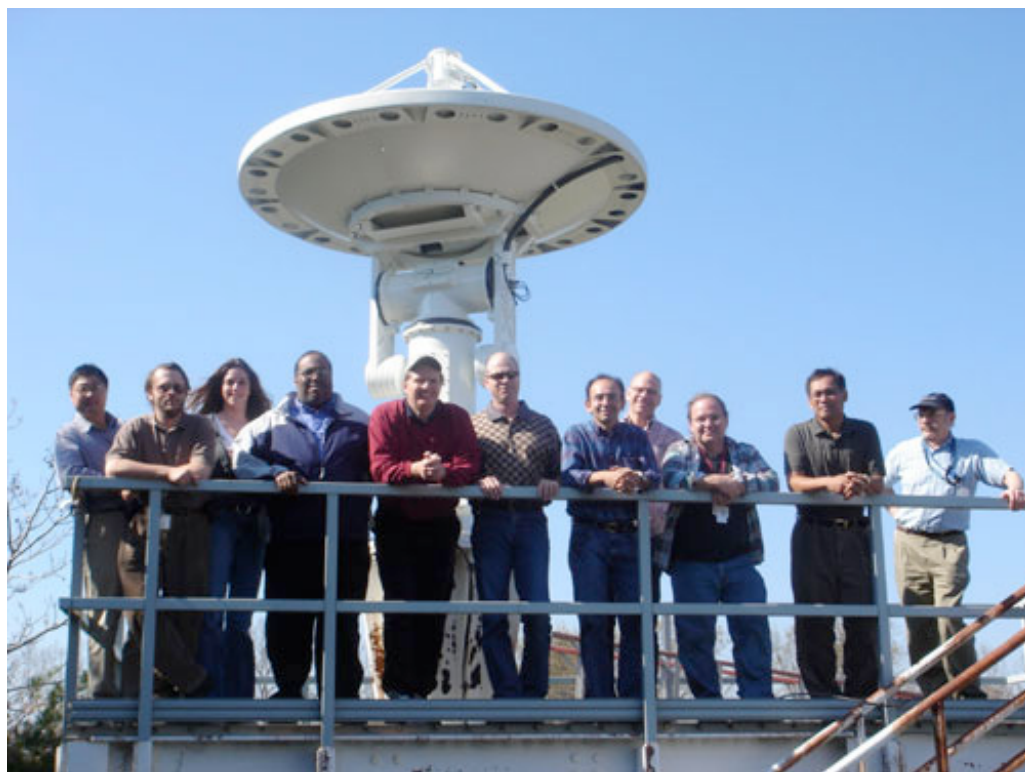

Figure 2: DRL staff shown on the mounting platform of the 2.4-meter diameter production antenna at Goddard. From left to right, the staff are Zhangshi Yin, Bob Bane, Nancy Dixon, Rodney Coleman, Stan Hilinski, Kelvin Brentzel, Pat Coronado, Bob Kannenberg, Glen Gardner, Dan Jacob, and Charlie Hoisington.

Providing data to the GES DISC is an important mission-related secondary activity of the DRL. The primary activity of the DRL will continue to be influencing cost-effective hardware designs, both on-board the satellite and the ground segment, and providing data system software tools, and the science instrument processing algorithms. Thus, this primary role of the DRL insures that a direct readout ground station will be able to cost-effectively acquire and process EOS and NPP/NPOESS satellite data for the benefit of real-time Earth remote sensing data users. The cost-effectiveness of direct readout ground systems has vastly improved in the last decade, with costs dropping from millions of dollars to thousands of dollars. The cost driver for any direct readout ground station continues to be the antenna. For an L-band system, the present cost is about $15,000, and for an X-band system, the cost is $120,000. The cost difference between the two systems derives from the more critical tolerances in fabrication of the more exacting surface for the higher-frequency antenna. The DRL and Goddard continue to work with the commercial sector to bring these costs down by providing data processing tools and algorithms that make the hardware more useful, which, in turn, creates a larger user base, which then leads to a larger production volume and, hence, lower costs. Thus, the bottom line is that the cost-effective services of the DRL help ensure that end users will continue to invest resources into having their own direct readout systems to acquire, process, and utilize Earth satellite data in near real time.

+ Privacy Policy and Important Notices
+ Sciences and Exploration Directorate
+ CISTO

Curator: Aimee Joshua
NASA Official: Phil Webster, CISTO Chief
Last Updated: Friday, 03-Aug-2007 11:46:02 EDT

CISTO News

- HOME

+ ARCHIVES

Editor
Mike Hollis

Associate Editor
Jarrett Cohen

Consultants
Lara Clemence
Jim Fischer
Nick Burke

NEWS UPDATES

Discover Cluster Expands with New Processors and Visualization Capabilities

By Jarrett Cohen

Once fully integrated, the new Scalable Unit (foreground) will add 1,024 processors to the Discover cluster.

To meet growing demand for supercomputing by NASA scientists and engineers, the NASA Center for Computational Sciences (NCCS) is adding a second 1,024-processor Scalable Unit to the "Discover" cluster. With this expansion, the Linux Networx Custom Supersystem will have a total of 2,560 processors and offer five times the computing power of the "Halem" HP AlphaServer SC45 system it replaces. NCCS and Linux Networx Inc. staff installed Scalable Unit 2 the week of July 23, and the NCCS expects to complete integration in August.

Discover went into production mode in January and has had usage rates over 90% on some days. Job sizes range from 4 (a single node) to 1,024 processors (256 nodes). To accommodate the augmented Discover's electricity and cooling needs, Halem was retired on May 1. The 1,392-processor Halem system served as the NCCS capacity computing platform for more than 4 1/2 years, supplying on the order of 50 million processor-hours to the NASA Earth and space sciences community.

NCCS and Software Integration and Visualization Office (SIVO) staff provided substantial assistance to move users from other systems, mostly Halem. "High-usage Halem investigator teams were each assigned

a coordinator to ease their transition," said Sadie Duffy, NCCS User Services Lead (CSC). Early on, all users received a User FAQ in their home directories; the FAQ and a User Guide are also available on the NCCS website. Some users found the move relatively simple, Duffy said, because they already computed on smaller Linux clusters in their departments. Familiarity also stems from Discover having nearly the same user environment as the NCCS "Explore" SGI Altix 3700 computer.

SIVO's Advanced Software Technology Group (ASTG) offered training for porting software codes to Discover. They also worked closely with scientists, especially those with high-use codes, to benchmark and tune their applications for improved performance and scalability, said Hamid Oloso, ASTG applications support lead (Northrop Grumman IT).

One of those high-use codes is ModelE, the newest atmospheric general circulation model at the Goddard Institute for Space Studies (GISS) in New York City. Results from ModelE simulations, mostly run at 4 by 4 1/2 degrees on Halem, were part of the U.S. contribution to the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (2007).GISS is currently testing and refining ModelE on Discover.

"The model used for our IPCC experiments runs roughly twice as fast on the new cluster compared to Halem without any changes in the code," said Reto Ruedy, GISS Climate Group Manager (Sigma Space Partners LLC). "Since shared-memory machines are becoming a rarity, we converted the parallelization of our code from OpenMP to MPI. This transition was made much less painful due to the support of Tom Clune and his group and the fact that we could take full advantage of the Earth System Modeling Framework (ESMF). This version of ModelE can simulate 100 years in 5 days, and we now are able to run our long experiments with a finer resolution."

On Discover, a 2- by 2 1/2-degree, MPI version of ModelE achieved even greater speed-up, as high as 3x per processor compared to Halem, noted ASTG Lead Tom Clune. Many other codes have seen 2x or better performance increases. In general, codes scale as well as on other NCCS computers, Clune said. Performance is particularly good on the Intel "Woodcrest" (2.66 GHz) processors inside the Scalable Units. While having a slower clock rate than the Base Unit's "Dempsey" (3.2 GHz) processors, they are 30 to 40% faster because of more efficient floating-point operations and speedier memory access, explained NCCS Lead Architect Dan Duffy (CSC). Scalable Unit 1 was formally accepted in April. Scalable Unit 2 will be slightly different in configuration, with 32 of its 256 nodes having 8 gigabytes of memory (versus the usual 4 gigabytes) for jobs that can benefit from it.

Teams whose applications produce large amounts of data often rely on scientific visualization to better understand their results. In a unique approach, the NCCS has fully integrated 16 Visualization Nodes into Discover. They can access all the same file systems as the rest of the cluster. These 4-processor nodes use AMD's "Opteron" (2.6 GHz) chips because they have superior memory-to-processor bandwidth, a capability particularly helpful for high input/output rates. Additionally, each Visualization Node has 8 gigabytes of memory and an NVIDIA Quadro FX 4500 high-end Graphics Processing Unit (GPU).

Linux Networx Inc. and NASA Center for Computational Sciences staff installed Scalable Unit 2 the week of July 23.

SIVO's Scientific Visualization Studio (SVS) will be working with users to define Discover's visualization environment. It will include common visualization software packages (e.g., IDL) and may offer both batch and interactive modes. "A growing area of interest is the use of GPUs for creating visualizations of simulations while they are running," said SVS programmer Jim Williams (GST, Inc.). "We hope to explore how we can use the Visualization Nodes to allow researchers to visually monitor their codes in real time. We may also be able to produce high-temporal-resolution animations that would require prohibitive amounts of storage were they to be done after the simulation finishes." Pioneer user access to the Visualization Nodes is available by contacting NCCS User Services at support@nccs.nasa.gov.

http://www.nccs.nasa.gov

http://sivo.gsfc.nasa.gov

+ Privacy Policy and Important Notices
+ Sciences and Exploration Directorate
+ CISTO

Curator: Aimee Joshua
NASA Official: Phil Webster, CISTO Chief
Last Updated: Friday, 03-Aug-2007 11:46:02 EDT

Information Science & Technology Colloquium Series

IS&T Award - 2007

Congratulations to *Patrick Coronado*, recipient of GSFC's Seventh Annual Excellence in Information Science and Technology Award

This award is presented annually to the Goddard employee(s) who best exhibit(s) broad, significant contributions to Goddard programs or projects in the areas of information science and technology. The award recognizes career achievement or a specific act of exceptional merit that was completed in the previous year.

Patrick Coronado,Senior Engineer and manager of the Direct Readout Laboratory in the Computational & Information Sciences & Technology Office

The 2007 award was presented to Patrick Coronado for his career achievements in satellite direct broadcast information technology since 1992. His most significant accomplishment during that time was leading the development of prototype software for capturing and processing direct broadcast data from the Terra and Aqua satellites. This software has been taken up by the private sector for incorporation into their systems, thus helping to enable a multi-million dollar industry. Patrick manages research and

development efforts in high data-rate satellite ground systems and information and multi-sensor data processing systems, including direct readout technologies for upcoming missions such as the NPOESS Preparatory Mission. He holds a Bachelor's degree from the University of Michigan in Mechanical Engineering and Mathematics, Masters degree in Aerospace Engineering – Space Systems, and a Masters in Engineering Management from The George Washington University..

The Spring 2007 Series concluded on May 16, 2007 with a special colloquium held in conjunction with presentation of the Center's seventh annual **Excellence in Information Science and Technology Award.**

Previous Winners: [2001](#) | [2002](#) | [2003](#) | [2004](#) | [2005](#) | [2006](#)

Information Science & Technology Colloquium Series
Responsible NASA Official: Paul Hunter
Curator: Patrick Healey
+ Privacy Policy and Important Notices
This file was last modified on Friday, 22-Jun-2007 17:00:11 EDT