GES DAAC Users' Working Group (UWG) Meeting Minutes
May 10-11, 2011

Notes:
- Meeting participants are listed at the end of these notes, as well as a partial list of acronyms used during these meetings.
- Suggestions and recommendations appear in red.

Day 1—May 10, 2011

Welcome (Steve Kempler, GES DISC Manager)

Logistics and Introductions…Agenda, Objectives, Roles and Responsibilities

- Steve reviewed the agenda and meeting objectives.

UWG Roles and Responsibilities (Eric Fetzer, JPL, UWG Chair)

Welcome to the GES DISC UWG

- Nearly 40,000 people used GES DISC data in fiscal year 2010, both domestically and internationally.
- The UWG and the Goddard Earth Sciences Data and Information Services Center (GES DISC) need to be aware of this community, both data providers and data users, whose needs the UWG represents.
- The goal of this meeting is to assess how and whether those needs are being met.
- The challenge before the UWG is to provide actionable recommendations to GES DISC and NASA HQ.

Welcome from Code 610 (Steve Wharton, Global Change Data Center [GCDC] Chief)

Welcome from Code 610

- The GCDC is part of the Earth sciences division.
- The mission of the GCDC is to build and operate data information systems in support of Earth science missions and research.
- Steve listed the components of the GCDC's approach.
- He pointed out that the GCDC includes the Global Change Master Directory (GCMD).

HQ Perspectives, HQ Role in the UWG, and HQ Mechanisms for Carrying out Recommendations (Ken Jucks, GES DISC Program Scientist, HQ)

HQ perspectives on NASA Earth Science Data Systems

- Ken explained his role as the GES DISC Program Scientist working within HQ.
- He discussed what the Distributed Active Archive Center (DAAC) is all about:
    - It provides a service to the science community—both good data and good analysis!
    - It facilitates the maximum science return from the various NASA data sets.
    - With a limited budget to work within, it needs to try to pursue the highest-value initiatives for the least cost.
    - The DAAC tries to be responsive to the changes—sometimes fast changes—in how science is done. If someone comes up with a new, truly useful way to use and display data, the DAAC tries to anticipate that the whole community may want to use it.

    - The DAAC provides a critical arm for all its mission projects. Space missions are a waste of money if people cannot use the data, and there is no point in launching rockets if NASA does not get the desired data as an end result.
- Ken listed this meeting's goals:
    - To help the DAAC do its job as well as possible
    - To help the DAAC know the areas where things are working well and where they may need to change
    - To act as a conduit from the community to the DAAC for ideas to improve the DAAC's service to the community
    - To maximize the science return from NASA's investments
- Ken laid out some of the rules for this meeting.
    - No idea is out of bounds
    - The effort is not to try to solve the DAAC world's problems overnight.
- He explained HQ's roles vis-à-vis the UWG and the DAAC.
    - Ken plays the role of advocate between the science community using the DAAC, the DAAC's funder (Martha Maiden), and the Program Scientists for the other DAACs.
    - His role is to listen and let himself be used as a sounding board for the best way to operate the DAAC.
    - Within her available budget, Martha balances the needs of all the data systems.
- Ken discussed mechanisms for carrying out recommendations and for reviewing and funding requests.
- He pointed out that the data centers do a lot of competitive, duplicative work resulting in duplicate or nearly duplicate data that need prioritization, reconciliation, or retention.

EOSDIS Perspective and Update (John Moses, Earth Science Data and Information System [ESDIS] Project)

ESDIS Project Report

- John provided an EOSDIS (EOS Data and Information System) perspective, starting with the concept that NASA is now in a golden age, with 12 operating satellites from which the data flow to users right away and relatively few lost satellites or missions. The science community is really getting the most out of these satellites right now.
- John presented a high-level overview of the various data centers, including the sorts of operations taking place in each as well as the locations where the various data are stored and where the processing of various levels of products/data (L1, L2, etc.) take place, functions that relate to where the expertise is located.
- He discussed the idea of duplicative storage of copies of data in multiple DAAC locations, a notion that might have been impossible at one time due to data storage difficulties but is possible today—if the need were compelling.
- It is not always clear or obvious where the data may be found, especially for outside users who cannot easily discern which data center (or even which organization, e.g., NOAA) stores the desired data or, where near-duplicate data exist, which data center stores the most current or complete version.
- With some data accumulating in orders of magnitude, the data storage infrastructure will need to be responsive to the increases in data volume.
- The effort of the DISC is to try to anticipate the next "tall pole." It used to be data storage and handling, which are no longer issues. Someone suggested that both provenance and data

integrity/quality are becoming the next issues to tackle. Specifically, when multiple data sets overlap, a user needs to understand which specific data are most appropriate for that user's purposes.

- John presented some EOSDIS performance measures for GES DISC and the rest of the data centers, including the amounts of data stored, the numbers of users and products, the rate of data growth, and the product distribution trends for each data center.
- He pointed out that technical infusion allows for and is critical to the continuing growth in data volume.
- He suggested providing a comment or review field in which a user could record possible integrity problems or tips for effective use, to enlighten other users.
- He presented web metrics, i.e., trends regarding unique and repeat visitors to each data center's web site. GES DISC has the third highest visitor rate out of all the DAACs.
- One attendee suggested that it would be very helpful to use the metrics analysis tool (NetInsight) to discover where authors of a published article found the data that they used and cited.
- Someone suggested there might be a way to survey those authors. This would be very helpful information.
- A proposed data citation policy would also be helpful but could not necessarily be enforced.
- The annual customer satisfaction survey indicates that GES DISC consistently rates higher than other Federal government organizations, higher than all industries surveyed, and higher than the rest of EOSDIS. In fact, in 2010 GES DISC jumped significantly to a rating of 80.
- The fact that NASA makes its data available to everyone in the world, transparent and free of charge, speaks loudly and reflects highly on the U.S. to what is becoming an increasingly international user community.

User Satisfaction Survey—ESDIS and DISC Project Reports (Jim Acker, GES DISC, Wyle)

DISC Project Report

- Jim reviewed some of what GES DISC did in 2010 to earn the jump in customer satisfaction
- This year, GES DISC added to the anonymous survey group those users who had contacted HelpDesk about using the GES DISC Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni).
- Customers lauded Giovanni and Mirador for their accessibility, ease of use, breadth of data, and powerful, intuitive search functionality, and more.
- Surveyed customers indicated that they would like the ability to access data at the pixel level and to download both spatial and temporal subsets at the same time.
- An area of constructive criticism involved the absence or inconsistent quality of documentation for various data from tool to tool or mission to mission. Some documentation was seen as very good, some inadequate. For example, a link to the documentation does not always take the user to the specific location of interest within the documentation but instead to the very beginning of the documentation.
- Nice-to-haves mentioned in the survey included data without haze and noise, the ability to construct the maximum possible longest data set with the highest possible resolution in binary and Network Common Data Format (NetCDF) along with the metadata, the desire for the least possible data latency, and the addition of a Castilian Spanish interface and data.
- The survey contained very few extremely negative comments.

State of the DISC (Steve Kempler)

State of the DISC

- Steve reviewed the GES DISC mission and summarized the expertise residing within GES DISC, including software engineering, science data management, mission support, and expertise on the operational active archive and distribution system with complete user services.
- He listed missions that GES DISC supports in the areas of atmospheric composition, atmospheric dynamics, modeling, hydrology modeling, precipitation, and Making Earth System data records for Use in Research Environments (MEaSUREs) data sets.
- He also listed multi-mission projects and technology projects initiatives that GES DISC supports.
- He provided an overview of data system evolution at GES DISC and the longevity of various data sets (e.g., TOMS, UARS, etc.) over the past 30 years.
- He covered activities that bring together heterogeneous data sets (subsetting) to address the special interests and needs of segments of the user community.
- He reviewed the 2011 GES DISC staffing and budget as well as the approximate anticipated budget and staffing through 2015.
- He listed a few significant 2011 GES DISC peer-reviewed publications, 2011 presentations delivered at conferences and workshops, invited Giovanni educational events, and a number of GES DISC-hosted seminar series.
- Steve described a new challenge for GES DISC: For example, TOMS data on the TOMS web site do not always match TOMS data on GES DISC.
- This documentation or communication issue needs to be addressed.
    - When there are different versions of data sets within the NASA archives, the distinctions and usages need to be explained on both/all web sites, which need to be cross-referenced. (Some users may not even be aware of the existence of differing versions of a data set.)
    - This also applies to user-created products that become popular with other users who may not understand the algorithms and parameters used to create the product, which can result in widespread publications whose logic is flawed. User-created products need to be accompanied by an explanation, but whose responsibility would this be, and how would it be done? (Someone pointed out that it is unrealistic to assume that users who create such products would take on this responsibility.)
    - Rama asked for ideas about how resolution of these problems might be achieved efficiently and accurately.
    - One difficulty relates to very old data sets, specifically how to systemize the creation of these data set caveats.
    - Someone suggested that there be an indication of the “certified” or sanctioned data set or standard data product. (This would also require that the non-sanctioned data be identified as such and be accompanied by a link to the sanctioned data.)
    - This would require a certification policy and process. Europeans have taken the lead on this, but only for future data, not archived data.
    - This issue really relates to good data stewardship.
    - Choreographing this effort would require collaboration between the DAAC level and the user level and a certification board and would require levels of certification.

- This is not a NASA-specific problem; someone suggested that GES DISC could look at how museums handle this, but others pointed out some problems that museums would not have faced.

Review of 2009 UWG Meeting Recommendations and Results (Steve Kempler)

10 recommendations

1. *Expand the UWG:* GES DISC added four new members to the UWG since 2009.
2. *Organize data by science discipline:* Since 2009, the parameter tree in Mirador has been organized along the lines of NASA's six Earth Science Focus Areas. (GES DISC is now looking for UWG feedback.) The DISC also initiated a Federated Search interoperability framework amongst Earth Science Information Partners (ESIP) Federation members that would enable Mirador to search other data providers.
3. *Keep an archive of user scenarios:* The User Scenario Archive suggestion provided a key part of the inspiration for the Earth Science Collaboratory (described later in these minutes). Steve indicated that GES DISC might not have done as much on this recommendation as could or should have been done.
4. *Increase access to multi-sensor observations:* Steve listed background information regarding support of this recommendation.
5. *Help further NASA's role in climate assessment:* GES DISC ties to modeling and data assimilation groups have strengthened with the ingest, archive, distribution of, and services for a number of data sets, functions that GES DISC is working to make more efficient. Also, discussion is in progress for potential relevant proposals.
6. *Enable transition for users in the applications and decision support communities:* Steve reviewed accomplishments since 2009 and then asked whether including an applications tab in Mirador would be a high-priority recommendation.
   - This might be especially useful for users who may not be intimately familiar with the data they are seeking.
   - Peggy brought up the SMAP data, which are ready to be made available to users but have not yet had a DAAC identified for storing these data. The delays involved in this effort, which affect potential users who cannot yet access the data, relate to this 2009 recommendation, not just regarding GES DISC but across all DAACs.
7. *Help ensure data quality and provenance:* Steve listed several successful GES DISC proposals in response to ACCESS (Advancing Collaborative Connections for Earth System Science) and AIST ESTO (Advanced Information Systems Technology Earth Science Technology Office) solicitations. GES DISC has also continued to expand Giovanni provenance. Steve asked whether there are other recommendations to pursue regarding this 2009 recommendation.
8. *Transition to support for future missions:* Steve listed known future missions that GES DISC will support as well as missions that GES DISC was recommended to service on which no decision has yet been made about which DAAC will service the mission.
   - Peggy asked how GES DISC should try to ensure consistency in multi-mission data sets.
9. *Allow subsetting and simplified formats:* These functions have been expanded, largely through deploying new Open-source Project for a Network Data Access Protocol (OPeNDAP) servers that support NetCDF download. In addition, the Simple Subset Wizard adds subsetting and reformatting services and makes existing services easier to use.

10. *Continue asking GES DISC UWG and science community for clarification of data service enhancements:* Steve indicated that this has not taken place on a wide scale since 2009. Steve asked for thoughts regarding whether GES DISC should engage the UWG in a quarterly teleconference to exchange information more regularly.
    - It was suggested that GES DISC and the UWG should meet semiannually (later in the meeting this was amended to every 15-18 months) and that Steve should send out a quarterly progress report to make sure he is keeping everyone abreast of ongoing efforts. In advance of each semiannual telecon, Steve will send out a note about topics of discussion.

Discussion Sessions:

Discussion Session—Lines of Business (LOBs) (Steve Kempler)

Lines of Business

- The LOBs represent the focus of GES DISC's thoughts as a data center
- Steve listed the LOBs such as technology services, interdata use, data brokering, measurement-based systems, data stewardship, data services, mission support (the DISC's most significant and time-consuming service), etc. He had hoped to take on a certified data archive (as discussed earlier), but its priority was too low to take on yet.
- Steve discussed the hurdles, the time investment, and the delays associated with a single NASA release and the need to find a faster and more efficient means of getting a release approved. (GES DISC percentages would be higher if this effort were not so time-consuming and cumbersome.) Someone suggested the possible solution of making the product belong to and reside at an entity other than NASA, allowing for the bypass or delay of the official NASA release. Others pointed out the difficulties this can introduce, e.g., NASA makes its data available freely and transparently, but another entity might not.
- Steve posed discussion questions about the LOB's appropriateness, consistency with user needs, and correct mix of efforts, as well as ideas for eliminating or adding LOBs and for pursuing LOB.
- A mindset should be in place such that, when PIs propose products, the DAAC needs to start right away with the development of appropriate and necessary documentation (provenance, source, etc.).
- Preparation of documentation could be incorporated into the release effort, such that checking for needed documentation would become one of the requirements for approval. The DAACs via ESDIS would need to establish expectations for the documentation and provide these to approving entities.
- Certification of data might come in different levels or values of certification (e.g., gold standard, silver standard, etc.).
- One user discussed the fact that downloading extremely large amounts of data can still take a lot of time, although performance speed is constantly improving. Chris suggested that using OPeNDAP might be a quicker, more direct means of getting the data, which the users do not really know about. GES DISC members indicated that they could help find efficient methods for users to achieve their ends when encountering problems such as this.

Discussion Session—New Services (Chris Lynnes, Chief Systems Engineer, GES DISC)

New Services

- Chris posed questions regarding what kinds of data discovery and access methods users find particularly useful and how can GES DISC obtain more input from the user community regarding access and discovery.
- GES DISC data discovery services:
  - Mirador development has been scaled back to a sustaining engineering level.
  - The Mirador external search (currently in test mode TS1) will allow users to request that the search be expanded to include external (Open Source) data sources. Thus far, GES DISC has not developed a means of identifying duplicates when, for example, internal and external sources include the same data. Chris anticipates that data from some external sources will be located very slowly, which raises the question of integrating the results to make them truly usable.
  - Usability issues include the need for the right balance between ease of use vs. introduction of new features. Chris wants to develop a Mirador Usability Sounding Board, to which GES DISC would send questions regarding usability quandaries encountered.
- Data services:
  - Chris examined statistics regarding which GES DISC services the users are accessing, currently and projected over the coming years.
- Data Quality Screening Service (DQSS)
  - DQSS is an ACCESS project that applies to AIRS data only, for now.
  - This requires establishing a "quality scheme," which can be very complicated; the scheme relies on identifying the levels of quality for a given data set *for each purpose*.
  - DQSS will take the low quality pixels and replace them with fill values, without destroying the original data or the applied mask.
  - Ultimately, this will allow the user to select the quality criteria for the search and include/exclude data based on their specific levels of quality.
  - Chris asked which user communities would most benefit from having this kind of screening service and which data sets should be prioritized for the rollout, after GES DISC completes MODIS Water Vapor data and Aerosols data and MLS Water Vapor data, which are already scheduled as the next candidates.
  - Someone suggested that OMI is a possible candidate.
  - Someone else suggested that GES DISC should work on everything for a given discipline, regardless of where the data are stored, and then move to another discipline. This suggestion drew consensus.
- OPeNDAP, Web Map Service (WMS), and Web Coverage Service (WCS)
  - OPeNDAP gives access at a very fine granularity level.
  - It presents Hierarchical Data Format (HDF) data as NetCDF/CF (Climate and Forecast) metadata, which enhances the tool's usability.
  - It also provides reformatting, including the ASCII and Network Common Data Format (NetCDF) formats.
  - Examples of potential users include industrial-strength scripters looking for subsets, "thick" client users, and internal systems such as Giovanni and the Simple Subset Wizard.
  - It provides a simple form to complete and a very quick response.
- Open Geospatial Consortium (OGC) Standards—WMS

    - A URL request returns a map image.
    - This was implemented with open-source MapServer.
    - Giovanni also supports WMS.
- OGC Standards—WCS
    - This returns "coverages" based on data variables in NetCDF/CF.
- Subsetting
    - There are semi-custom tools for some products.
    - The process usually involves HDF in and HDF out.
    - This was implemented as REST* URLs.
    - It involves subsetting at the time of the download.
    - The subsets are implemented as the user requests come in.
    - GES DISC has about 100 subsettable data sets including, for example, specific AIRS, MLS, TOMS, OMI, and TRMM data, and models.
- Format conversion
    - Some custom code for some L3 and L2 data sets exists, in response to user requests.
    - This involves HDF → NetCDF/CF.
    - This improves the usability of tools.
    - GES DISC is moving toward external tools where possible, such as OPeNDAP and "LATS for Dummies" (Lats4d) (based on the Grid Analysis and Display System [GrADS]).
    - In response to a question, Chris pointed out that GES DISC needs to look into whether it should/will modify the metadata on spatial subsets.
- Simple Subset Wizard, an ESDIS-sponsored effort
    - For subsetting, users would like to say, "Just give me the data from time 1 to time 2 for this spatial box."
    - The current approach is to search for granules, view the granules, subset the granules, select the subset option, re-enter the special box, and so on—a very cumbersome process.
    - The Simple Subset Wizard will be available for GES DISC and nine other data centers.
    - It has a very simple interface in which the user selects the data, then enter keywords and dates for the desired data range.
- Giovanni evolution
    - The original goal was to allow scientists to concentrate on the science, not bungle through use of the technology.
    - Giovanni has become somewhat expensive to maintain and difficult to add new features.
    - There have been increasing requests for L2 and point data, which Giovanni does not currently handle.
    - The G3 architecture was started but never completed; some unforeseen factors stalled the project but eventually led to a move to the new G4 focus.
    - Moving forward, GES DISC will implement new projects using Agile Giovanni (G4), which will include the intended G3 improvements and will address the need for L2 and point data.
- Community Giovanni

- GES DISC has built a prototype for discovering data and storing them in a knowledge base that will be publicly available: Community Giovanni.
- Community Giovanni will provide an interface for the addition of metadata, such as the owner of the data.
- It will allow users external to the GES DISC to support their own data sets and portals.
- It is being designed so that it can be used with ECHO as well.

- Chris listed "roads not taken," i.e., efforts GES DISC did *not* pursue (yet), as well as the efforts the DISC pursued instead.
- Someone suggested extending education efforts, not specifically for the use of Giovanni but rather for the use of the data.
- Someone else suggested that GES DISC should make Giovanni available for smart phones and iPads.

Technology Initiatives Discussion Sessions—ESDIS Evolution Initiatives:

Discussion Session—Land, Atmosphere Near-real time Capability for EOS (LANCE) (Bruce Vollmer, Data Support, GES DISC and Karen Michaels, ESDIS)

LANCE

- This LANCE effort is about making available to users near-real time data with minimal latency, i.e., within three hours of data collection.
- Data originate from the Terra (MODIS), Aqua (MODIS, AMSR-E, and AIRS), and Aura (OMI and MLS) spacecraft.
- The latency for Terra is better than for Aqua and Aura.
- LANCE data are provided via FTP or by subscription.
- Data products are freely available to registered users.
- Each element of LANCE currently provides user support through existing resources.
- Present application areas include hurricanes, volcanoes, fires, oil spills, dust storms, air quality, snow/ice, and weather.
- In August 2010, GES DISC added a feature so that users can self-register.

Discussion Session—User Registration System (URS) (Bruce Vollmer and Andrew Mitchell, ESDIS)

User Registration

- GES DISC is trying to build a central URS but not to eliminate existing systems, at least not in the near term.
- GES DISC has made an effort to gather input from internal and external users.
- The URS will enable improved collection of metrics on users and usage of resources to help evaluate system enhancements and the impact of new tools and services.
- Metadata collected will include user registration, user authentication, profile and password expiration and renewal, profile maintenance, etc.
- The URS will provide substantial benefits to EOSDIS (improved metrics collection, etc.) and to the user community.
- The proposed URS solution includes an online browser-based GUI that will allow users to perform self-registration; in addition, the core system will be augmented, for example, to support "tiered users" and to provide a mechanism for resolving user registration conflicts across external systems.

- GES DISC plans a phased approach for transition.

Discussion Session—REVERB/ECHO (Andrew Mitchell, ESDIS, and Chris Lynnes)

REVERB/ECHO

- The EOS Clearing House (ECHO) is a middleware layer between Earth Observing System (EOS) data and science data users.
- ECHO currently contains about 95 million granules, with all but two of the data centers participating so far.
- REVERB (which means re-echo), the new client replacing the Warehouse Inventory Search Tool (WIST), is intended to improve the user experience regarding user satisfaction and performance.
- ECHO has significantly improved search performance, ingest performance, and more.

Discussion Session—Coherent Web (John Moses, ESDIS, and Long Pham, GES DISC)

Coherent Web

- The next steps in this effort include creating a coherent web presence.
- Currently the many NASA web sites have a variety of different looks, each of which is very well liked by its customers. The goal is to maintain the web sites' flexibility and responsiveness to their respective user communities and still move toward a coherent, more consistent web presence.
- The user interfaces are separate from the underlying web functionality.
- The Coherent Web team has mocked up a general layout. The intent is to make all the web sites and all the services available from a given web page.

<u>Technology Initiatives Discussion Sessions—Community-Based Initiatives:</u>

Earth Science Collaboratory and Other Technical Initiatives (Chris Lynnes)

Community Based Initiatives

- The situation today is that Earth science data, services, tools, and analysis methods are not easy to use.
- The proposal is for the convergent evolution of all data centers to an Earth Science Collaboratory (ESC), a rich data analysis environment that would provide access across a wide spectrum of Earth science data, provide a diverse set of services, and more
- The ESC would include a tool library divided into four sections: a privatized section, a community section, a contributed section, and a personal section. It will allow for sharing of tools and discussion of tool usage.
- The ESC would also include a data library, a workflow library, and a laboratory notebook, all divided into the same four categories listed above and available for data sharing and discussion.
- The laboratory notebook would include articles and the sources/data used to research the articles, the algorithms used, the analysis approach, etc.
- The corresponding information across these libraries would be cross connected.
- The idea is that a user could duplicate the same approach, data, etc., that another user used for researching a published article. This would enable extended research based on an original approach by allowing users to recreate and further analyze data that were previously used.

- Someone asked how best to compel users to record this sort of information. Someone suggested that available metrics might allow GES DISC to capture and display testimonials from users who have had success or benefited from the ESC; testimonials from highly reputed scientists would encourage other users to share their own experiences.
- Someone commented that this looks like a framework that will evolve over time and might become a sort of social application.
- In the long view (the next 7+ years), GES DISC wants to get community consensus, expand the architectural concepts to full architecture, establish incentives to fit into the ESC architecture, generalize proofs of concept to robust versatile services, etc.
- Regarding the definition of ESC standards, someone suggested that GES DISC might look at how some of the major science collaboration efforts put their science together, e.g., aerosols assessment and national climate assessment, IPCC, and CMIP5.

[The remainder of this discussion with Chris took place on Day 2.]

- Chris listed Earth science user communities, from which clusters could be formed, i.e., a cluster can be formed from an informal work group. Clusters are how things are accomplished in ESIP.
- He covered a number of initiatives that relate to specific segments of the user community.
    - He discussed a "why so difficult" rant that took place at a 2008 Earth Science Data System Working Groups (ESDSWG) meeting (notes available online by searching for "why so difficult" and "NASA"), which led to steps taken in the area of ESIP Discovery Clusters.
    - He asked whether GES DISC should push harder in this area. It takes a lot of effort to move clusters, so pushing harder would need to add enough value to be worth it.
- He discussed the ESDSWG Technology Infusion Working Group, which was mandated as part of ACCESS Project wins.
    - He mentioned the Working Group on Information Systems and Services (WGISS) Architecture Development Committee, which co-authored a white paper on "Challenges to Seamless Data Discovery and Access to Remote Sensing Data" that laid out problems but no solutions. Chris asked the UWG for ideas about what should come next and possible solutions.
    - He mentioned the WGISS Atmospheric Composition Portal but pointed out that the effort needs more science direction.
    - Someone asked who pays for these efforts—NASA does.
- Chris posed the questions about whether GES DISC has the right amount, not enough, or too much involvement in technology-oriented communities and in science communities (to which a user typically need an invitation).
    - Someone pointed out that these two communities—technology and science—really go hand in hand and likened Chris's role to a matchmaker between the two.

GES DAAC User Working Group (UWG) Meeting Minutes
May 10-11, 2011

Notes:
- Meeting participants are listed at the end of these notes, as well as a partial list of acronyms used during these meetings.
- Suggestions and recommendations appear in italics.

Day 2—May 11, 2011

Follow-up on Day 1 Discussions (Steve Kempler)
Steve pointed out that Day 1 discussions revealed that some UWG members were just learning about existing techniques at this meeting. He asked what GES DISC needs to do to ensure that non-UWG users can learn about these techniques as well.
- Someone suggested that, with new interns arriving soon, maybe GES DISC could offer a short tutorial for them.
- Someone suggested creating online tutorials, demos, and/or short videos. Since these consume users' time, they should be very brief and pointed and should list the duration of each so that users will know the time commitment beforehand.
- Actionable item: GES DISC should have some sort of brief user email that gets sent out infrequently, introducing people to new capabilities. GES DISC would need to set this up in its privacy policy and would need to create mailing lists such as a DAAC-wide email group. GES DISC could also send out reminders of existing features directed at new users. GES DISC might assemble a "cookbook," i.e., step-by-step instructions; users could contribute their own ideas for modifications they have made to their own version of the instructions to customize or improve upon it, which would be useful for new users.

Technology Initiatives Discussion Sessions—GES DISC Evolution Initiatives:

Discussion Session—Cloud Computing (Long Pham, GES DISC)
Cloud Computing
- As the DAAC grows, so do the cost of maintaining the DAAC and the footprint regarding servers, which has led to the introduction of cloud computing.
- Cloud computing can be an application that runs off your web browser or an infrastructure as a service and can result in a drastic reduction in processing time.
- The nebula cloud prototype used AIRS data and algorithm. The preliminary result was a significant reduction in the time spent processing data, from 50 hours down to 10 hours.
- GES DISC still needs to work out kinks in the cloud, such as input and output (I/O) of data, to get a better understanding using the cloud more broadly.
- The actual data results were identical when using the cloud, other than metadata reflecting the time spent processing the data.
- Cloud computing is cheaper in terms of storage, but I/O will be very costly, both for input and again for output.

- Someone warned that what is currently freeware may eventually cost money (as experienced by JPL); in addition, users have no control over when the server for the freeware is taken down for maintenance.
- There is great interest in cloud computing, despite these concerns.
- Someone pointed out that there is an economy of scale, i.e., the more people who use it, the cheaper it becomes per user. In addition, in the event that freeware begins to cost money, nothing would prevent GES DISC from moving to different cloud software.

Discussion Session—Web Redesign (Rashid Chowdhury, DISC Web Task Lead, GES DISC)

Web Redesign

- The GES DISC home page is cluttered and does not represent the big picture of what the DISC is all about.
- In the next six weeks to two months, GES DISC will be integrating a "top hat" on the home page, in accordance with the direction being recommended by the Coherent Web project.
- The goal is to unclutter the home page and eliminate the need to scroll.
- Someone suggested that, where the top hat displays an acronym, it would be helpful for users if hovering over the acronym were to reveal a full explanation or translation of the acronym.

Discussion Session—Multi-sensor Data Synergy Advisor (MDSA) and Provenance (Greg Leptoukh, GES DISC

MDSA, provenance

- Greg discussed the dilemma of how a user cannot find the desired data, and if he can, cannot access it, and if he can, does not know how sound the data is, and so on.
- Because it is so easy to get to data, GES DISC needs to provide clear information about the data to help resolve this problem.
- Provenance is very poorly handled by almost everyone, in large part because of the varying provenance between varying data, that is, how to compare and reconcile these varying levels of provenance. For example, data from MODIS Terra and Aqua Aerosols do not agree.
- MDSA is about providing to users information that may not be obvious regarding provenance and other issues related to data quality.
- Greg covered some of the challenges in dealing with data quality:
    - Different users, e.g., data providers and data recipients, perceive quality differently.
    - There are many different qualitative and quantitative aspects of quality.
    - Methodologies for dealing with data quality are just emerging.
- Regarding quality control and quality assessment, the former reflects that the data conform to the expectations of the algorithm, while the latter takes place after the fact, based on data analysis of "fitness for the purpose."
- Working with data quality has been very difficult, much more difficult than anticipated.
- Greg asked to what extent the data centers should pursue these activities and how knowledge from one group should make it to other communities.
    - Users might find this a very valuable service, especially users who are unable to reconcile variations in provenance.
    - Someone clarified that this is more about repairing error propagation than improving data quality. The latter insinuates that this work is about optional value-adds when really the work should be seen as required error corrections.

Discussion Session—Aerostat and Validation (Greg Leptoukh)

Aerostat, validation

- Aerostat will allow for the improved correlation of data from many satellites for the purpose of effective data fusion.
- This is a big step because it eliminates some user work currently associated with this effort.
- Greg pointed out that this is building a collaborative environment and asked for suggestions about how it should be done.

<u>Mission Support Discussion Sessions:</u>

Discussion Session—Multi-Sensor Initiatives (Greg Leptoukh)

Multi-Sensor Initiatives

- Greg listed a variety of existing multi-sensor activities, such as certain ACCESS and Research Opportunities in Space and Earth Sciences (ROSES) proposals.
- GES DISC is an expert at multi-sensor initiatives, e.g., bringing together different disciplines and gaining expertise in working with multiple formats.
- Greg asked whether some of GES DISC's core funding should go into these activities (i.e., divert funding from other areas.
- Someone asked whether successes in these areas result in GES DISC taking on more work, pulling people like Greg away from core activities. The group discussed the possibly negative results and the impacts on already stretched staff members.

Discussion Session—Model Data (Peter Smith, DISC Science Support, GES DISC)

Model Data

- The new data set will have three grid types: observed minus analysis, observed minus forecast, and observed values.
- The planned user interface will provide a normal level of user support.
- Peter asked whether this data set fits into GES DISC's climate modeling support role.
- Peter also asked whether GES DISC should consider ingesting other data sets of this type to facilitate modeling research studies.

Discussion Session—MEaSUREs Projects Data Support (Bruce Vollmer)

MEaSUREs Projects Data Support Report

- Bruce provided a status report for the MEaSUREs projects, first mentioned at the 2009 UWG meeting, and laid out the timeline for these projects.
- He described MEaSUREs support and touched on some highlights of the MEaSUREs projects
- A MEaSUREs portal is available through the main GES DISC web page.

Discussion Session—Sustaining Completed Projects (Steve Kempler)

Sustaining completed projects

- This refers to the need to address data services life cycles. Projects are funded to create or do something, and then funding stops when the project is completed. This results in a lack of funding for ongoing maintenance, bug or breakage fixes, or fulfillment of user requests for modifications or enhancements.

- Assuming these unfunded services have value in the eyes of the UWG, Steve asked for ideas about what GES DISC and the UWG might do to address this.
- Someone wanted to get an idea of the relative cost of each of these unfunded services, to help understand the cost vs. value of each unfunded item. GES DISC has not yet identified the relative and cumulative costs. This will require estimating the cost reductions that might result from fulfilling each unfunded service, i.e., a cost-benefit analysis.
- Someone suggested that GES DISC might have to require users with requests to explain and justify each request, which would put the onus on the requesting user to prove need. In other words, GES DISC would not write a proposal to respond to the request but rather the requesting party would write it.
- Someone expressed the concern that, if GES DISC users do not get what they want without having to write a proposal, they might look elsewhere (i.e., another data center), resulting in reduced usage of the DISC's offerings such that the DISC and its offerings could eventually become obsolete.

Discussion Session—Applications Activities Related to the USDA and DSS (Bill Teng, GES DISC)

Applications activities (USDA, DSS)

- An application project may involve customized data, production of new data, new tools, etc.
- Bill briefly discussed two examples of applications projects and reviewed work that GES DISC has done so far in this area.
- He pointed out that these applications are intended to run without maintenance or ongoing human involvement.

Presentations we did not have time to present:

Data Stewardship and Data Provenance activities – Steve K and Greg

Our new proposed data citation policy (generate strawman) – Chris

Executive Session: Discuss Findings, Suggestions, Recommendations (UWG, ESDIS)

- The UWG felt this session was very relevant, helpful, and valuable. There was no obvious was a conflict between the direction that GES DISC and the UWG see things going; the GES DISC presentations more or less reflect the direction the UWG wants to go.
- Recommendation: *Help the GES DISC provide consistent, high quality documentation for its user community.* This was prompted by a bullet in Jim Acker's overview of the User Satisfaction Survey "An area of constructive criticism involved the absence or inconsistent quality of documentation for various data from tool to tool or mission to mission. Some documentation was seen as very good, some inadequate." This is a major issue in flight projects. Remember, we are serving a community of nearly 40,000 data users. However, this problem goes beyond the GES DISC to those projects, but Headquarters can help ensure documentation quality.
- Recommendation: *Improve tracking of productions history and provenance for the rapidly increasing number of data sets in the GES DISC archives*. Some physical quantities are produced by a variety of sensors (aerosol, for example), while MEaSUREs and other projects

are combining those observations. How does the GES DISC provide information about the history of these data? Documentation is clearly part of the solution, but others need to be enacted.

- Recommendation: *Establish a mechanism to increase collaboration between the GES DISC and its users community*. The goal here would be stronger interactions between science needs and interests and the services GES DISC provides.
- Recommendation: *Establish a user registration process.* The user metrics acquired through registration would allow the DISC to provide great value moving forward; for example, GES DISC could discover the number of users. (GES DISC would need to assure users that there would be no spam associated with registration.) There may be privacy issues involved that need to be investigated before running ahead with this.
- Related recommendation: *Explore the option to go in as either a registered user or an unregistered user.* (This did not appear to be a recommendation widely endorsed by the UWG.)
- Related recommendation: *GES DISC needs email distribution lists, based on knowledge gleaned from the user registration process.*
- Recommendation: *Create the role of official DISC project scientist.* The role would include responsibility for gathering the kind of user input and feedback that is currently gathered ad hoc by members of the DISC, to help advise on decisions made within GES DISC. This scientist would be a liaison with the science community. (Steve explained the history of GES DISC's past attempts to install a project scientist.) The issue of a formalized loop between users and the mission project scientists needs to be resolved.
- Related recommendation: *Use NASA Energy- and Water- Cycle Study (NEWS) to test this role.* NEWS involves global data sets and has the potential for interaction. Related to that, there are international efforts (e.g., the Global Energy and Water Cycle Experiment, or GEWEX) to place similar data on the same grid.
- Recommendation: *GES DISC needs an improved marketing strategy for explaining what is happening with the tools, data, etc., to keep users updated.*
- Recommendation: *Establish a user database.* As user and usage metrics accumulate, GES DISC should cross-reference the certified data and the registered users i.e., more analysis of the nature of the user community. This also relates to data permanence, which would allow GES DISC to enhance the value of data certification by making the certification "stamp" persistent. Since creation of a user database will likely be costly, it should follow a set of specific goals, like data certification.
- Recommendation: *GES DISC should ensure that users always have access to the latest, most current data updates.* This relates to data certification and the Earth science collaboratory or similar data portals.
- Recommendation: A *mechanism will be needed for heritage data*. Data documentation, provenance and persistence are all implied in this recommendation.
- Recommendation: *Once a user registration system is in place, GES DISC should refer to user statistics to identify the heavy users and consider asking them to join the UWG.* GES DISC asked whether the present makeup of the UWG is an adequate representative cross section of the user community, but no one knew for sure since they do not yet know the makeup and needs of the GES DISC user community.

- Recommendation: *GES DISC should let user metrics inform priority decisions as the metrics begin to accumulate.* The UWG needs more information about the routine core activities for each mission as opposed to broad, cumulative core information.
- Recommendation: *Reach out to a broader community*. The UWG wants to see NASA more energetically infusing its data and tools into the high school and college levels to continue creating new generations of users. Outreach to the Applications community was also mentioned at the first UWG meeting in April 2009.
- Recommendation: *Include higher level functions in Giovanni such as computing climatologies and anomalies so that users do not have to download the entire data set.* [Note: Steve Kempler mentioned users are becoming more interested in Level 2 data from Giovanni. See his comment in the Lines of Business presentation on p. 8 of the meeting notes.]
- Recommendation: T*he next UWG meeting should be in 15 to 18 months, with an intervening telecon or two prior to that*. GES DISC should consider coordinating the timing of its UWG meetings to coincide with ASDC UWG meetings.

Partial List of Acronyms Used During This Meeting

Acronym	Definition
AIST	Association for Iron & Steel Technology
ACCESS	Advanced Cosmic ray Composition Experiment for the Space Station
CF	Climate and Forecast
DAAC	Distributed Active Archive Center
DIF	Data Interchange Format
DISC	Data and Information Services Center
DQSS	Data Quality Screening Service
ECHO	EOS Clearing House
EDOS	EOS Data and Operations System
EOSDIS	Earth Observing System Data and Information System
ESDIS	Earth Science Data and Information System Project
ESDR	Earth Science Data Records
ESDSWG	Earth Science Data System Working Groups
ESIP	Earth Science Information Partners
ESTO	Earth Science Technology Office
GCDC	Global Change Data Center
GCMD	Global Change Master Directory
GES DISC	Goddard Earth Sciences Data and Information Services Center
GEWEX	Global Energy and Water Cycle Experiment
GIOVANNI	GES DISC Interactive Online Visualization ANd aNalysis Infrastructure
GOZCARDS	Global OZone Chemistry And Related trace gas Data records for the Stratosphere
GRaDS	Grid Analysis and Display System
HDF	Hierarchical Data Format
LANCE	Land, Atmosphere Near-real time Capability for EOS
LATS	Library of AMIP Data Transmissions Standards
LATS4d	LATS for Dummies
LOBs	Lines of Business
MDSA	Multi-sensor Data Synergy Advisor
MEaSUREs	Making Earth System data records for Use in Research Environments
NetCDF	Network Common Data Format
NEWS	NASA Energy- and Water- Cycle Study
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
ROSES	Research Opportunities in Space and Earth Sciences
SIPS	Science Investigator-led Processing Systems
URS	User Registration System
UWG	User Working Group
WCS	Web Coverage Service
WGISS	Working Group on Information Systems and Services
WIST	Warehouse Inventory Search Tool
WMS	Web Map Service

Attendance List
GES DAAC User Working Group (UWG) Meeting
May 10-11, 2011

Name of Participant		Organization	Role
Last Name	**First Name**		
Adler	Robert	U. of Maryland	UWG Member
Albuyrak	Arif	GES DISC	Science Algorithm Development
Alcott	Gary	GES DISC	GES DISC Operations
Ahmad	Juraiya	GES DISC	Senior Scientist
Berrick	Stephen	NASA HQ	HQ
Bosilovich	Michael	NASA/GSFC/GMAO	UWG Member
Bruton	Diane	GES DISC (ADNET)	GES DISC—Tech. Writer
Chen	Aijun	610.2 /GMU	GES DISC
Chowdhury	Rashid	610.2/ADNET	DISC Web Task Lead
Ding	Feng	610.2	GES DISC Science Support
Fetzer	Eric	JPL	UWG Chair
Froidevaux	Lucien	JPL	UWG Member
Hearty	Thomas	610.2	GES DISC Science Support
Hedge	Mahabal	GES DISC (ADNET)	GES DISC SE Team
Hua	Hook	JPL	UWG Member
Issa	Suehail	GES DISC (ADNET)	DISC Group Manager
Jucks	Ken	NASA HQ	DAAC Program Scientist
Kempler	Steven	GES DISC	GES DISC Manager
Khayat	Mo	GES DISC (Wyle)	DISC Deputy Group Manager
Kirschbaum	Dahlia	GSFC/NASA	UWG Member
Kusterer	John	NASA/LaRC	ASDC Manager
Leptoukh	Greg	610.2/NASA	GES DISC
Lindsay	Francis	423/NASA	Internation/Interoperability
Lynnes	Chris	GES DISC	Chief Systems Engineer
Massie	Steve	NCAR	UWG Member
Mitchell	Andrew	423/NASA	ESDIS/ECHO/URS
Moluar	Gyula I.	NASA GSFC	UWG Member
Moses	John	NASA GSFC/ESDIS	DAAC Engineer
O'Neill	Peggy	NASA GSFC	UWG Member
Ostrenga	Dana	GES DISC (ADNET)	DISC Staff
Parker	Lindsey	NASA/Langley	Operations/User Services
Pham	Long	GES DISC	Infrastructure Support
Ramapriyam	H.K. "Rama"	GSFC ESDIS	ESDIS Project Manager
Savtchenko	Audrey	610.2/Wyle IS	MEaSUREs/ ALOS
Smith	Peter M.	GES DISC	DISC Science, MERRA
Strub	Richard	610.2	Mirador
Teng	Bill	GES DISC (Wyle)	Science Support
Vollmer	Bruce	GES DISC	Data Support

Name of Participant		Organization	Role
Last Name	**First Name**		
Wade	Gail	GES DISC	Security/Desktop Support
Walter	Jeff	NASA/GSFC/ESDIS	Deputy Project Manager/Tech.
Wei	Jennifer	610.2	GES DISC Science Support
Wharton	Stephen	GCDC	Parent organization
Youn	Wainie	GES DISC	Events Support