**NASA Science Mission Directorate**
**Research Opportunities in Space and Earth Sciences – 2009**
**Advancing Collaborative Connections for Earth System Science**
**(NNH09ZDA001N-ACCESS)**

The ACCESS Cooperative Agreement Notice (CAN) solicits projects that provide strategic, near-term improvements in NASA's Earth science data and information systems by leveraging existing technologies. The specific objectives of the 2009 ACCESS announcement are (1) to enhance the discoverability and availability of existing data system tools and services from NASA, other federal agencies, academia, and the private sector; and (2) to help users better use Earth science data by enhancing connections between the data and information on their quality, best practices, community input, and other associated information.

A total of 35 proposals were received for this announcement. All proposals were peer evaluated using a combination of mail and expert panel review. The Earth Science Division of NASA's Science Mission Directorate selected 11 proposals for two-year awards pending satisfactory budget and work plan negotiations. Some of these awards will be partially funded. These projects will help to further improve NASA's Earth Science Division's heterogeneous and distributed data and information systems.

---

---

**Edward Armstrong/Jet Propulsion Laboratory**
**Improving Discovery for Coastal Marine Web Services and Resources**

Many resources available for coastal ocean research and management remain underutilized. Typically, the emphasis in the past has been on increasing access and usability of remote sensing satellite products from NASA data centers. Significant progress has been made in this regard although access and discovery mechanisms still remain disjointed. Less attention has been paid to discovery and usability to ocean insitu records and circulation model products, because typically these are organized and maintained on a smaller regional level such as a university or smaller division of a larger national agency. Thus, in many cases web services from these regional activities suffer from underutilization.
In this effort we will focus on improving discovery of these regional coastal ocean web services and data portals, including databases for satellite imagery, in situ and field measurements, and GIS coverages as a few examples. In addition, we believe that ocean model product web services remain underutilized, including regional ocean models. In the first phase of our effort we intend to build an online catalog of these resources that will allow users to quickly discovery the availability of web services and data for their region of interest, physical parameter of interest or specific regional project of interest, or any combination of these. In the second phase of our effort we will integrate those services that allow callable data retrieval syntax such as WCS/WMS, OPeNDAP, JPL Datacasting into a client within the framework of the EASy GIS, the most advanced GIS for marine research and data products. Currently it supports over 40 NASA remote sensing products and a growing list of in situ and ocean model products. Using a web

interface (client) to the core software a user will be able to integrate, view, overlay and interrogate products that have been linked to our discovery catalog.

---

**Andrew Bingham/Jet Propulsion Laboratory**
**Web Feed Discovery and Integration Service**

Consumers of Earth science data are increasingly using multiple geophysical parameters and related georeferenced information to perform their investigations. For example, researchers studying the migration patterns of marine animals are interested in both the physical and biological processes within the ocean and therefore utilize sea surface temperature and ocean color data together. They might also use other sources of data and information, such as in-situ observations, photographs, maps, news articles, and even eyewitness reports to aid their investigations. RSS (Really Simple Syndication) feed extensions designed to share georeferenced information (GeoRSS) and data (DataGeoRSS) provide a key for integrating information and data.

We will deploy a Web Feed Discovery and Integration (WFDI) service that provides a registry for cataloging and automatically discovering GeoRSS and DataGeoRSS feeds. We will build technologies to find GeoRSS feeds on the web and scraper technologies to create DataGeoRSS feeds from online data holdings. The WFDI Client will provide the ability to i) discover feeds through its interface to the feed registry, ii) subscribe to feeds, iii) filter feeds to find specific items within a feed, iv) visualize the georeferenced information and data on a map and v) enable cross-feed correlation. This latter capability is a powerful technique that exploits commonalities across disparate feeds to combine data sets in meaningful ways and thereby enable the discovery of new information.

---

**Frances Boler/UNAVCO**
**Discovery and Delivery of Space Geodetic Data Products from Distributed Archives**

Space geodetic science and other disciplines using geodetic products have benefited immensely from open sharing of data and metadata from global and regional archives. The International GNSS Service (IGS) is an example of coordinated sharing of data to enable the production of orbit and other products that benefit nearly all civilian users of GNSS data. Ten years ago, two of the IGS Archive Centers, Scripps Orbit and Permanent Array Center (SOPAC) and the NASA Crustal Dynamics Data Information System (CDDIS) collaborated with UNAVCO to create the GPS Seamless Archive Centers (GSAC) in an effort to further enable research with the expanding collections of GNSS data then becoming available. The GSAC partners share metadata to facilitate data discovery and mining across participating archives and distribution of data to users. This effort was pioneering, but was built on technology that has now been rendered obsolete. As the number of geodetic observing technologies has expanded, the variety of data and data products has grown dramatically, exposing limitations in data product sharing. We propose to expand the original GSAC capability for multiple geodetic observation types and to simultaneously modernize the underlying technology by implementing Web Services at CDDIS, the UNAVCO Data Center, and SOPAC. A science partner

(University of Nevada, Reno) will test these Web Services by incorporating them into their daily GNSS data processing scheme. The effort will include new methods for quality control of current and legacy data that will be a product of the analysis/testing phase performed by UNR. This proposed development will contribute to an improved datum, the International Terrestrial Reference Frame, which is required for the accurate detection and monitoring of sea level change, the distribution of terrestrial water and atmospheric water vapor and other global-change related processes.

---

**Michael Goodman/ NASA Marshall Space Flight Center**
**Instant Karma: Applying a Proven Provenance Tool to NASA's AMSR-E Data Production Stream**

Current procedures for capturing and disseminating provenance, or data product lineage, are limited in both what is captured and how it is disseminated to the science community. This proposal brings together a team of NASA and university researchers with expertise in NASA Earth science data systems, science algorithm development, and provenance collection/dissemination. This team, led by Michael Goodman, Deputy Manager for Earth Science Office at MSFC and manager of the AMSR-E Science Investigator-led Processing System, proposes to apply a proven provenance tool to the generation of NASA's AMSR-E standard products, with an initial focus on sea ice products. The team will utilize Karma, a provenance tool developed by Co-I Beth Plale at Indiana University. Karma is a provenance collection and representation tool designed and developed for data driven workflows such as those used to generate NASA standard products. Other team members include Dr. Thorsten Markus, Head of the Cryospheric Sciences Branch at NASA/GSFC, member of the AMSR-E science team and responsible for the science algorithms used to generate the AMSR-E Sea Ice products; and Dr. Rahul Ramachandran and Helen Conover at University of Alabama in Huntsville. Both have rich experience with metadata issues and Earth science software applications, and have been involved in NASA standards committees and working groups.

This project will integrate Karma into the AMSR-E SIPS production environment. The AMSR-E SIPS generates Level 2 and Level 3 data products from AMSR-E observations, which are key data sets for research in both the Climate Variability and Change and Water and Energy Cycle focus areas. An initial focus on Sea Ice processing will allow the project to engage the Sea Ice science team and user community in customizing Karma for NASA science data. Metadata gathered by Karma will be presented to the AMSR-E data community via an interactive web application.

---

**Hook Hua/Jet Propulsion Laboratory**
**Tracking Production Legacy of Multi-Sensor Merged Climate Data Records**

Multi-decadal climate data records are critical to studying climate variability and change. These often also require merging data from multiple instruments such as those from NASA's A-Train that contain measurements covering a wide range of atmospheric conditions and phenomena. The science Co-I of this proposal is also recently funded under a MEaSUREs NRA to provide a merged and multi-decadal climate data record of

water vapor measurements from sensors on A-Train, operational weather, and other satellites. The data sets are being assembled from existing data sources, or produced from well-established methods published in peer-reviewed literature. However, the immense volume and inhomogeneity of data often requires an "exploratory computing" approach to product generation where data is processed in a variety of different ways with varying algorithms, parameters, and code changes until an acceptable intermediate product is generated. This process is repeated until a desirable final merged product can be generated. Typically the production legacy is often lost due to the complexity of processing steps that were tried along the way. The data product information associated with source data, processing methods, parameters used, intermediate product outputs, and associated materials are often hidden in each of the trials and scattered throughout the processing system(s). We propose to help users better interpret the exploratory process of this production legacy by enabling the tracking of data, metadata, associated materials, algorithms, and parameter changes used during the production of these merged and multi-sensor data products. By leveraging existing provenance tools, we will capture the metadata associated with the exploratory computing and present the data product provenance back to the users. We will also develop generic multi-platform clients to be plugged into existing code to communicate production information back to the provenance collection tool. To improve data knowledge and use, we will also develop a web portal enabling users to track and visualize the product and processing information collected by the provenance tool. For any product generated, a fully traceable processing lineage will be available that includes the production methods, parameters, source data, and associated information used. This capability will enable one to cite data products in science literature with links to its full data and service provenance.

Water vapor and cloud observations from the current generation of NASA sensors, especially those on the A-Train, cover a wide range of atmospheric scales and a wide class of phenomena. Our proposed system will improve knowledge of NASA's Earth science data quality and production legacy of multi-sensor and multi-decadal water vapor data records. The size, heterogeneity and complexity of global-scale and long-term climate change often requires more complex data processing whose production legacy must be tracked and preserved for traceability and scientific justification.

We will integrate Web Service-base tools for multi-platform data provenance tracking (such as the "Karma Tool for Provenance Collection and Storage") into existing data production environments. Generating a merged multi-decadal climate data record of water vapor measurements requires potentially different processing environments. We will develop generic client plugins in Matlab, IDL, Python, and C/C++ that can be easily integrated into existing code and will communicate via Web Services to the provenance collection tool. To improve data knowledge and use, we will also develop a provenance web portal enabling users to track and visualize the product and processing information of the generated data products.

---

**Gregory Leptoukh/NASA Goddard Space Flight Center**
**AeroStat: An Online Platform for the Statistical Intercomparison of Aerosols**

The largest uncertainty in computer models of the Earth's future climate is the magnitude of the primary and secondary aerosol climate feedbacks. In order to understand the contribution of changing tropospheric aerosol burdens on climate and constrain the associated uncertainties, it is imperative to quantify the current global distribution of aerosols. However, differences between near-coincident satellite observations of aerosol parameters are often larger than the respective reported uncertainties. The scientific community is currently concentrating on identifying, understanding and resolving specific differences between tropospheric aerosol observations.

One cannot address these issues without a fairly comprehensive environment in which to perform the data intercomparisons. Such a platform must provide not only the data itself, but also ready access to correlative information on data quality and data lineage (provenance).

We propose to create and implement AeroStat, an online environment for the direct statistical intercomparison of global aerosol parameters in which the provenance and data quality can be readily accessed by scientists. AeroStat builds upon the framework of an existing NASA online data visualization and analysis tool, Aerosol Giovanni (http://giovanni.gsfc.nasa.gov/), and leverages other ongoing projects (TRL 6-9) to provide the most cost-effective means of comparing satellite, ground-based and model aerosol data.

AeroStat will provide a creative, collaborative research environment where users can seamlessly share AeroStat workflow execution, algorithms, best practices, known errors and other pertinent information with the science community so other users can reproduce their results AeroStat’s collaborative environment will allow users to share the full details of individual case studies by registering them on a web-accessible Wiki.

---

**Christopher Lynnes/NASA Goddard Space Flight Center**
**Data Quality Screening Service**

NASA's Remote Sensing data typically include a rich set of quality information. These can range from simple quality flags to complicated bitmasks to external criteria such as cloud cover. Additional quality information is often included in external documentation or peer-reviewed literature and may include latitude dependencies, physiographic criteria (e.g., not valid over deserts), temporal functions, etc. The explanations of these quality factors also appear in a variety of forms, such as data product documentation and journal articles. Thus, it is a laborious process for data users to locate the quality explanations, read the quality indicators and code an interpretation of the indicators. The effort expended is compounded many times, because each user must repeat the process him or herself. If it is merely difficult for science users, it is virtually impossible for machine applications to assimilate and use the quality information; they must either reject the data product or ignore the quality information. We propose to solve this problem through the

construction of an ontology-driven quality screening service. The service will be deployable as a simple REST or SOAP Web Service at data centers, taking as input the original data product and producing as output the same product, but with quality screening applied as requested by the user or client. In addition, predefined quality screens will be provided according to the cognizant science team's recommendations as stated in their documentation.

The quality screening indicators and recommendations will be encoded in a community-based data quality ontology, leveraged from an ongoing NASA project, which will allow it to be reused, expanded and maintained as new missions (such as the Decadal Survey missions) come online. In addition, the ontology-based approach will make the quality screening and associated information usable by machine-based applications such as models and decision support systems.

---

**Charles Meertens/UNAVCO**
**Science Data Systems for Satellite and Airborne LiDAR Data**

The large volumes of data and derivative products produced by the forthcoming NASA DESDynI and ICESat-2 satellite missions will stretch the limits of the data management and processing capabilities of existing Earth science data systems. Access tools for DESDynI data will need to address the diverse requirements of the scientific community who will use these data to study deformation of the solid Earth and cryosphere and vegetation structure. They will also need to serve the application community's requirements in equally diverse areas ranging from earthquakes and glacier surges to wildfires and deforestation. For these hazards applications there is a requirement for timely, easy access to higher level, low latency, products in common data formats via simple to navigate web-interfaces. Recent Cyberinfrastructure developments, when integrated into existing Earth science data systems, can provide a framework for distributed data access and enhanced processing capability. To develop this capability ahead of the DESDynI and ICESat-2 missions, we propose to implement a system that enhances access to existing LiDAR data sources hosted at the National Snow and Ice Data Center DAAC, Goddard Space Flight Center LVIS Data Center, UNAVCO, and the San Diego Supercomputing Center (SDSC). Satellite LiDAR data from ICEsat, high altitude airborne LiDAR data from LVIS, and low altitude airborne LiDAR from the UNAVCO and SDSC's extensive Earthscope catalog will be incorporated into the OpenTopography Portal. This system provides integrated access to high-resolution LiDAR point data and on-demand processing capability for user-specified topographic data products. New web services will connect the NASA data archives at GSFC/LVIS and NSIDC to the OpenTopography portal and the other datasets available there. Integration of these distributed data archives via an easy to use web portal will enhance exposure of these data and significantly streamline user access.

---

**Jerry Pan/Oak Ridge National Laboratory**

**Delivering NASA Earth Observing System (EOS) Data with Digital Content Repository Technology: A Software System to Promote Best Practice in Digital Provenance and Effective Access of Content and Associated Metadata**

Digital content, including Earth Science observations and model output, is an essential part of contemporary scientific research activities. Not only is the rate of archiving for such content increasing rapidly, but there is also an increase in derived and on-demand data product creation and consumption. As a result of these trends, scientific digital content has become even more heterogeneous in format and more distributed across the Internet. In turn, this makes the content more difficult for providers to manage and preserve and for users to locate, understand, and consume. Specifically, it is increasingly harder to deliver relevant metadata and data processing lineage information along with the actual content, particularly when there are multiple ways of delivering the content, including the increasing use of web services. Readme files, data quality information, production provenance, and other descriptive metadata are often separated in the storage level as well in the data search and retrieval interfaces available to a user. Critical archival metadata, such as auditing trails and integrity checks, are often even more difficult for users to access, if they exist at all.

We propose to address these challenges by using and extending the capabilities of a contemporary digital object repository to work for science data and metadata delivery. Digital repository technology has been used for digital libraries at great success, and we believe it can also be applied to the more complex needs of Earth Science data management. We will demonstrate this capability in the context of an existing modeling and synthesis data center project for the North American Carbon Program (NACP) as the primary science context and one of the more complex data projects for the ORNL Distributed Active Archive Center for Biogeochemical Dynamics (ORNL DAAC) as a second context.

There are three high-level objectives in this project:
1. Demonstrate the applicability of a digital object repository technology to science data. Based on our preliminary work, we expect to couple the Fedora Repository and a Drupal-based Graphic User Interface (GUI) as key elements of a next-generation NASA Earth system science data center infrastructure, using datasets collected as part of the NACP Modeling and Synthesis Thematic Data Center (MAST-DC) as the primary science context.

2. Use this implementation to enable better and more consistent access to critical metadata, including processing lineage information and administrative metadata, using the capabilities inherent in a digital repository (multiple streams for a given object and remote data streams). The enhanced metadata access ensures that science digital content becomes more transparent to the end user, with provenance and quality control information readily available.

3. Demonstrate how data providers can more easily and effectively manage science data sets, associated metadata, processing lineage, and quality control/data provenance information. A consistent process, with associated user interfaces, application programming interfaces (APIs) can be used by the data provider to ingest, update, and modify a dataset for metadata changes or additional content dissemination revenues. Particularly in the context of the ORNL DAAC data, this work will demonstrate potential technology migration paths for existing data operations.

In addition, the successful completion of this project will provide a foundation for the improved long-term preservation of NASA Earth Science data using open standards, which should prove useful in the ongoing development of EOSDIS and the delivery of data for current and future missions.

---

**Mark Parsons/University of Colorado**
**ASIS: Aggregation of Services for Ice Sheets**

Ice sheet mass balance is a critical - the critical - parameter for forecasting sea level in the coming decades. While ocean thermal expansion and mountain glacier ice loss currently account for most of the rate of sea level change, sea level changes will be increasingly dominated by changes in the ice sheets, and these changes will increase significantly over this century (in fact, this is happening already). Moreover, changes observed in the ice sheets within the past decade have been larger, faster, and more variable than previously forecast in many areas. The problem of glacier and ice sheet response to changing climate and ocean conditions is more complex than was anticipated. Many research efforts are currently underway, gathering data in support of better assessments and modeling.

Ice sheet mass balance may be determined in several ways: by direct measurement of volume change; direct measurement of mass change; or by an accounting of net inputs and outputs. All the approaches are required to give the best assessment of rates and causes, and each requires management of extensive and distributed data sets. For example, the “accounting” method requires an assessment of snow accumulation rates, ice thickness, ice flow, and basal melt rate. These parameters are measured by a myriad of separate geophysical methods (field work, satellite, model results), spanning many different efforts by many projects and PIs. Data sets may be point measurements, traverse (line) measurements, or spatial grids of modeled/interpolated data. Moreover, the data may vary significantly over time, in particular, for ice flow, and in some areas ice thickness. Understanding, making use, and even just discovering this broad range of data is a fundamental challenge for polar researchers.

The informatics and scientific community have been developing a variety of services to help assess, interpret, and analyze these diverse data. These services (many developed by NASA) include mapping and visualization, automatic data subsetting or aggregation, format conversion, change notification, and so on. However, currently there is no ready way to discover all these services. Current approaches to the issues of data and service discovery, data publishing, and data annotation often rely on centralized mechanisms such as metadata directories or catalogs and data centers that work with investigators to

produce and publish formal data and documentation. Conversely, data providers increasingly self-publish their data but often in an ad hoc fashion and without the levels of service and support that formal archives can provide. The result of this is that users often do not have broad knowledge of the services that are available for their data of interest.

Kevin Werbach recently discussed the reasons for the phenomenal success of dotcom companies eBay, Amazon, and Google, noting "all of them aggressively open up their technical interfaces, allowing other sites to plug into them, or projecting themselves out to the rest of the Web ... the new paradigm ... is syndication. Open up your core assets and turn them into a platform; don't hide behind high walls and expect everyone to come to you (http://werblog.com/2006/02/syndication-in-action/)." Following these successful models, this project seeks to "aggressively open up" the technical interfaces to polar science data and web services based on those data by implementing several new applications of existing, open-source technologies to create a working prototype of a federated and community-enabled approach to data and service sharing and discovery. The project will focus on data services important to understanding ice sheet mass balance from both field and remote sensing measurements. However, our approach is generalizable and extensible to other disciplines and other applications. This will help sustain long-term usability and evolution of the tools and services we develop.

---

**Brian Wilson/Jet Propulsion Laboratory**
**Lightweight Advertising and Scalable Discovery of Services, Datasets, and Events Using Feedcasts and Social Tagging**

Discovery and use of Web Services for querying, accessing, and processing Earth Science datasets is hampered by the lack of an open and web-scalable services and data collection registry that provides rich search for all available services and datasets, interesting geophysical events, and data granules relevant to studying those events. Existing registries, like GCMD, ECHO, or GEOSS, provide limited search capabilities (only text keyword), often have difficult interfaces for registering services or datasets, each use their own metadata standard, don’t evolve their metadata fields (registered information) rapidly to suit user needs, and compete with each other for adoption, thereby fragmenting the user base. To find all available web services, the user must search all three registries, and even then the results may omit many unregistered services and datasets.

The PI has promulgated a "service casting" approach in which services are openly advertised on the web in Atom syndication feeds, which are searchable at Google and can also be aggregated to provide smart, faceted search (semantics beyond text keywords). The serv-cast v1 (metadata) standard is currently being vetted by ESIP Federation members and a catalog of published servcasts is already being accumulated. Servcasts can be used to search by service taxonomy, look up the service interface (e.g. WSDL), machine auto-invoke the service, or click through to service documentation for humans. If a provider's services change or expand, the advertisement can be modified and re-cast.

The service provider is in control, and can add metadata fields, or even information intended for its own use, to the extensible servcasts at any time.

We propose to develop a lightweight service advertisement mechanism that will improve web discovery of provider's Earth Science services, while also being backward-compatible by providing search over existing metadata repositories. The objective is to provide a one-stop search box in the browser where users can search by keyword and service taxonomy. By combining service casting with meta-search, we will provide a novel and complete solution:

1. Provide a lightweight, easily-authored mechanism for ES service producers to publish and promote their services and datasets;
2. Enable service consumers to easily find and invoke services, both those published as service casts and those registered in the GCMD, ECHO, and GEOSS repositories ("meta-search");
3. Provide a rich search interface as a browser plug-in that has modern, web 2.0 capabilities: search term suggestions for the user, term synonyms and broadening using semantics, search by user tags, and integrated social tagging so users themselves can enhance the categorization of services, rank them by usefulness or performance, tag them for use in collaborative service chains, etc.

Similarly, metadata describing interesting geophysical events (hurricane tracks) or periodic structures (El Nino), with links to related datasets and services, can be advertised in discoverable event-casts. Under this work, we will define the event-cast standard, leveraging ISO metadata standards, and populate a body of Hurricane casts as examples. Users will be able to tag services and events using a pre-defined hierarchical taxonomy or their own categories, publish new events, link datasets to events, and rank services by suitability for purpose, performance, availability, etc. The social tagging and semantic search capabilities will be implemented by extending Noesis technologies developed by Co-I Ramachandran.

Service casting and meta-search will improve the 'findability' and usability of all of the service and dataset metadata currently registered in existing repositories: GCMD, ECHO, and GEOSS. Increasing the use of Web Services will increase the use of NASA's Earth science datasets, especially for large-scale automated workflows that support climate science.

---