

ADVANCING COLLABORATIVE CONNECTIONS FOR EARTH SYSTEM SCIENCE (ACCESS)

SMD ROSES – NNH06ZDA001N
Program Element A.19

Award Announcement – October 30, 2006

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. This ACCESS solicitation called for proposals to develop web-based, science-user clients that utilize NASA's EOS Clearing House (ECHO) middleware technologies. These clients are intended to exploit the underlying capabilities of ECHO and serve as functional prototypes demonstrating the ease of development and reuse of ECHO clients by science and technology professionals.

A total of 14 proposals were received for this announcement. All proposals were peer evaluated using a panel review. The Earth Science Division of NASA's Science Mission Directorate selected 2 proposals for one-year awards pending satisfactory budget and work plan negotiations. These projects will help NASA further improve existing Earth Science Division's heterogeneous and distributed data and information systems.

Wilson, Brian Jet Propulsion Laboratory
AQUA: Automated Data Query & Access for Large Earth Science Datasets

A key challenge facing climate scientists is the difficulty of "scaling up" their statistical analysis to cover time periods of years to decades. If one wants to compare model grids from one or more (e.g., atmosphere) models to L2 or L3 retrievals (of temperature, water vapor, aerosol optical depth, etc.) from multiple EOS instruments, or just inter-compare the instruments, millions of EOS granules need to be located and then staged onto disk in order to perform the analysis. And inevitably, the process must be repeated as the models or comparison algorithms are refined. Currently, the data must be "ordered" (staged onto disk at the DAAC's) by a human filling out a web form; order sizes are usually limited to a week or two at a time; and the response to the user comes via email. The EOS Clearinghouse (ECHO) provides services for space & time query, order entry, and automated delivery of granules via FTP push or pull. (Each of these service requests is forwarded to the appropriate data provider.) We propose to develop an ECHO client that will take the human out of the loop and enable transparent, machine-to-machine, automated data query and access to multiple EOS datasets on a large scale.

To accomplish this, we will develop a set of machine-callable Web (SOAP) Services on top of ECHO data query & order entry services that together will automate the following multi-step process (or Use Case):

1. Query ECHO for datasets (collections) that contain the physical variables of interest,

2. Query those datasets for granule ID's that satisfy the desired space/time constraints,
3. Locate granules already on-line at the DAAC's or in the user's file cache,
4. or Order via ECHO their staging onto disk,
5. Fetch the granules using FTP URL's (or have them pushed),
6. or Access variables from the data files in-place using OpenDAP URL's to subset,
7. Analyze the data by calling the scientist's data fusion algorithm,
8. Repeat for progressive time "chunks" until the desired multi-year period is covered.

Each of our client services (query, locate, order, fetch, etc.) will be a composite service, which automates and hides the complexity of the multiple ECHO (SOAP) calls required to accomplish the task. Each service will have a simple SOAP interface described in standard WSDL, and be published (callable) at multiple web servers. Once these services are available, they will be assembled into an automated workflow to perform the desired scientific analysis (steps 1-8). The Order/Fetch/Analyze/Repeat service cycle will automatically adapt the size of the time "chunk" to the disk space available for staging data (at the DAAC's and client site). Using future Grid virtualization, the storage & compute resources required for a particular analysis job might be discovered and allocated on the fly, and paid for later on a utility bill.

Currently, a scientist wonders: Why can't I just push a button to (space/time) query, locate on-line, and/or automatically stage onto disk a year of AIRS, MODIS, and MISR L2 or L3 products and then use them in an automated, repeatable science analysis? In the near future, this will be possible and she will only have to worry about the monthly bill.

Although the primary goal of this project is to develop automated machine-to-machine services, we will leverage existing SciFlo capabilities, and use the AJAX programming paradigm, to layer on top of the SOAP services a dynamic web browser interface for human users. The user will specify a simple "chain" of services in the dynamic GUI and then the desired workflow will execute automatically.

Houser, Paul Institute of Global Environment and Society, Inc.
A Water and Energy Cycle EOS Clearing House Client

Water has long been known to be essential to human activity. Both plants and animals upon which humans depend, require water to survive. The ability to monitor and to study the amount and location of water is critical for the management of earth's resources. To this end, metadata in NASA's EOS Clearing House (ECHO) is an essential information resource for connecting water users to the best available data. In response to the NASA-

ROSES A.19 Advancing Collaborative Connections for Earth System Science (ACCESS) call to enhance and improve existing components of the data and information systems infrastructure supporting NASA's Earth science research communities, we propose to develop a new Water and Energy Cycle (WEC) EOS Clearing HOuse (ECHO) client (WECHO). The purpose of this client is to use ECHO middleware technology to provide access to and discovery of data resources related to the NASA Global Water and Energy Cycle Focus Area (one of 6 such NASA research focus areas). Science and technology professionals will be able to access NASA data and services via a web-based tool tightly integrated with ECHO Application Program Interfaces (APIs) to combine and then display via the client web browser data sets and subsets within the realm of ECHOs' holdings. Our objective is to establish practical applications that benefit the research of the community of known science users and extend those benefits to society at-large.

Interfacing with the ECHO API, the users of this new ACCESS client will be able to browse NASA's water and energy cycle related holdings and then display characteristics of the data set(s). This new portal will incorporate statistical and quantitative graphical displays allowing the user to apply educated human insight and analysis in the formulation of new hypothesis and the validation of current theories.

The portal will be based upon industry standards for performing web-based computing. The proposed portal will also offer some command-line based functionality for use in automating general tasks.