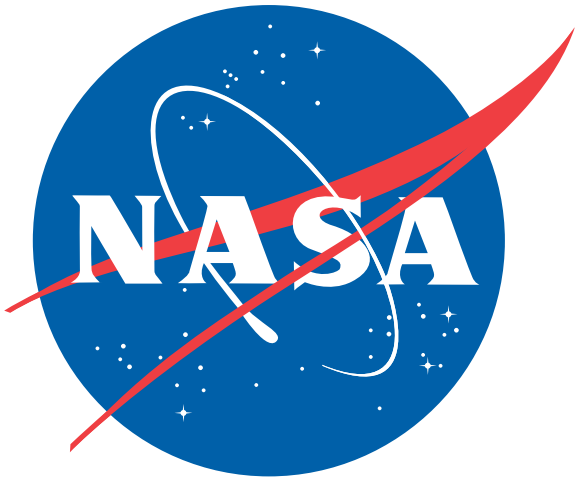

National Aeronautics and Space Administration

May 5, 2004

NNG04ZY4001N

MINI-SOLICITATION FOR MASTER NASA RESEARCH ANNOUNCEMENT NNG04ZY4000N

Earth Science Technology Office
Advanced Information Systems Technology (AIST)

Data Mining and Discovery Technology Demonstrations

Proposal Deadline: July 6, 2004

Master
NASA Research Announcement
NNG04ZY4000N
Advanced Information Systems Technology (AIST)

NASA Research Announcement
Mini-Solicitation
Soliciting
Data Mining and Discovery Technology Demonstrations
for
Period Ending
July 6, 2004

NNG04ZY4001N

Issued May 5, 2004

Earth Science Technology Office
Goddard Space Flight Center
Greenbelt, MD 20771

The NASA vision is: *To improve life here*
To extend life to there
To find life beyond

The NASA mission is: *To understand and protect our home planet*
To explore the Universe and search for life
To inspire the next generation of explorers

…as only NASA can.

The Earth Science Enterprise (ESE) is one of six NASA enterprises seeking to fulfill the agency's vision and carry out its mission (http://www.earth.nasa.gov/visions/index.html). The ESE mission is to understand and protect our home planet by using our view from space to study the Earth system and improve predictions of Earth system change. The ESE, working with its domestic and international partners, provides accurate, objective scientific data and analyses to advance our understanding of Earth system processes and to help policy makers and citizens achieve economic growth and effective, responsible stewardship of Earth's resources. The ESE research program aims to acquire deeper scientific understanding of the components of the Earth system, their interactions, and the consequences of changes in the Earth system for life. These interactions occur on a continuum of spatial and temporal scales ranging from short-term weather to long-term climate and motions of the solid Earth, and from local and regional to global.

The frontier of Earth system science is to: (1) explore interactions among the major components of the Earth system – continents, oceans, atmosphere, ice, and life, (2) distinguish natural from human-induced causes of change, and (3) understand and predict the consequences of change. NASA has established six scientific focus areas for these complex processes. These scientific focus areas are: Atmospheric Composition, Carbon Cycle and Ecosystems, Climate Variability and Change, Earth Surface and Interior, Water and Energy Cycle, and Weather. Roadmaps have been developed to summarize the technology, observations, modeling, field campaigns, basic research, and partnerships needed over time to achieve the long-term goals for each of these focus areas (http://earth.nasa.gov/roadmaps/). The roadmaps for the Carbon Cycle and Ecosystems, Climate Variability and Change, and Weather focus areas provide the strategic framework for research under this NRA mini-solicitation. ESE focus areas are interrelated and must eventually be integrated to arrive at a fully interactive and realistic Earth system representation.

Five fundamental questions drive ESE research:

- How is the global Earth system changing?
- What are the primary causes of change in the Earth system?
- How does the Earth system respond to natural and human-induced changes?
- What are the consequences of change in the Earth system for human civilization?
- How will the Earth system change in the future?

In accordance with the Earth Science Technology Office (ESTO) Master NASA Research Announcement (NNG04ZY4000N), the ESTO Advanced Information Systems Technology (AIST) Program announces this mini-solicitation for proposals addressing data mining and discovery technology demonstrations. This mini-solicitation provides additional details governing the proposed activities that supersede the general guidelines announced in the Master NRA. Proposals are to be submitted electronically via the AIST Master NRA proposal submission e-book accessible via the ESTO home page at http://esto.nasa.gov/. A printed, signed postal submission is also required, and both are due by the date and time specified in section V of this mini-solicitation. The postal submission is the official submission method for determining the received date and time for the proposal.

I. AIST Master NRA Mini-solicitation Topic: Data Mining and Discovery Technology Demonstrations

This mini-solicitation requests data mining and discovery technology demonstrations that address technology objectives selected from section I.1 and that address at least one ESE science challenge listed in section I.2. Proposals that address more than one ESE data type from NASA archives (indicated in the data sources for the selected challenge) will be given additional consideration.

Proposals will be evaluated using the criteria listed in the Master NRA Appendix A. The proposal's discussion of relevancy to ESE includes addressing the value of the proposed application to meet the science challenge and in answering key science questions in the ESE strategic research plan. References are provided in AIST Master NRA information available at http://esto.nasa.gov/. The appropriate selection of data suitable for responding to a well-founded data mining science question will be evaluated as part of the relevancy criteria. The proposal's technical approach must include an assessment of the potential for the proposed technology to be used for challenges beyond that selected by the proposer for technology demonstration.

I.1 Technology Objectives

Data mining technologies are sought to facilitate knowledge discovery in support of ESE science and application goals, by enabling automated approaches to data searching and delivery of data products on demand. With complex, heterogeneous data stored in distributed petabyte archives, an open system approach to enable discovery, access and delivery of Earth science data and services is needed. Included are data and service locator technologies leveraging the web and commercial approaches, tailored to the unique demands of the geo-spatial Earth science data sets. Researchers are invited to propose in the following subtopics:

1. Tools and support for warehousing, data mining and knowledge discovery

   Tools are needed for seamless multidisciplinary access to and operation on massive, distributed archives of heterogeneous data, models, and data processing algorithms. Better machine support is needed to automate hypothesis generation, change detection and algorithm development. Achieving machine-assisted knowledge

discovery via proper content-based metadata and semantic data search is of particular interest.

2. Technologies to facilitate physical and logical data queries and/or access mechanisms for multi-disciplinary Earth science data

   Large, long-term Earth science data sets must be readily accessible to users of all Earth science disciplines to support the development of accurate Earth system models. Data organization and management techniques and query representations to enable search, navigation, exploration and discovery are among the technologies sought in this area.

3. Techniques to facilitate customized application-oriented data services

   Techniques such as data subsetting, reformatting, and others can improve the analysis of scientific data and have the potential to contribute significantly to knowledge and information discovery.

I.2 **ESE Science Challenge**

I.2.1 **Ocean Biology and Biogeochemistry Data Mining**

NASA's ESE has collections of data available that contribute to the ESE Research focus areas objectives on climate variability and change, as well as carbon cycling and ecosystems. The physical climate and biogeochemical (carbon) cycles are linked within a complex Earth system. The data available in the ESE can be utilized for research into various questions that increase our understanding of the Earth System, allowing coastal managers to better protect natural resources and related commerce. Examples of exploratory questions one might ask that a data-mining tool provide data to answer include: "How can sea surface temperature data or optical properties, such as turbidity flags, be used to predict the onset of a coral bleaching event or examine overall coral health?" or "What is the frequency of harmful algal blooms within the Gulf of Mexico?" The outcome of individual research projects should include a product to enable coastal managers to better protect natural resources and commerce. Some examples include a predictive map of a marine area that may be subject to a coral bleaching event within a two-day notice period, or a capability to detect the onset of a harmful algal bloom.

Data Sources for ocean biology and biogeochemistry

Ocean Color (Level 3 gridded global data):
Mapped files - ftp://samoa.gsfc.nasa.gov/pub/sdps/V4/L3SMI/
Binned files - ftp://samoa.gsfc.nasa.gov/pub/sdps/V4/L3BIN/

Coastal Zone Color Scanner:
http://daac.gsfc.nasa.gov/data/dataset/CZCS/index.html

Ocean Color and Temperature Scanner (OCTS)/ Sea-Viewing Wide Field-of-view Scanner (SeaWiFS)/ Moderate Resolution Imaging Spectrometer (MODIS) Terra/ MODIS Aqua – Goddard Space Flight Center (GSFC) Distributed Active Archive Center (DAAC)
http://daac.gsfc.nasa.gov/data/dataset/OCTS/index.html

SeaWiFS
http://daac.gsfc.nasa.gov/data/dataset/SEAWIFS/index.html

MODIS Terra – GSFC DAAC
http://acdisx.gsfc.nasa.gov/data/dataset/MODIS/index.html

MODIS Aqua – GSFC DAAC
http://daac.gsfc.nasa.gov/data/dataset/MODIS-Aqua/03_Ocean/index.html

Subsets of all these products are available online via ftp through the GSFC DAAC Data Pools - http://daac.gsfc.nasa.gov/data/datapool/index.html

National Oceanographic and Atmospheric Administration (NOAA)/ National Oceanographic Data Center (NODC) (in situ data)
http://www.nodc.noaa.gov/General/getdata.html
http://www.nodc.noaa.gov/col/projects/access/seabass.html

Jet Propulsion Laboratory's Airborne Visible/Infrared imaging Spectrometer (AVIRIS) data (aircraft, and all other NASA and NOAA and Navy aircraft Light-Detection and Ranging (LIDAR) data)
http://aviris.jpl.nasa.gov/html/data.html

SeaWiFS Bio-optical Archive and Storage System (SEABASS) in situ bio-optical archive data
http://seabass.gsfc.nasa.gov

I.2.2 **Data Mining for Climate and Weather Models**

I.2.2.a Climate variability and agriculture

NASA ESE research links satellite data with the earth science challenge of predicting the droughts and floods associated with the El Niño Southern Oscillation (ENSO). The objective is to provide regional decision-support systems with useful tools for agriculture resource management. Specific goals are to analyze remote sensing data from the new generation of NASA Earth Observing System (EOS) sensors and to link ENSO projections from global climate models (GCMs) and regional climate models (RCMs) to agricultural projections. The layers of information derived from the remote sensing and climate information are then integrated into a regional Geographic Information System (GIS).

Data from the most recent NASA instruments (including MODIS and Tropical Rainfall Measuring Mission (TRMM)) are used to monitor drought and flood through vegetation state

and its temporal and spatial variability, and precipitation respectively. Analysis aims at detecting the signal of large-scale effects such as El Niño, as well as vegetation responses to intra-annual variability of key parameters like precipitation and surface temperature.

Information concerning key parameters such as land cover type, vegetation condition and surface temperature are monitored using NASA MODIS data. The combination of high spatial and spectral resolution with the high signal-to-noise ratio of MODIS data results in accurate and reliable vegetation indices that are calculated using input from several bands. The derived products include individual band spectral reflectance information in the visible and near-infrared wavelengths, and MODIS Vegetation Indices (Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)). The indices are analyzed in the context of change detection to identify areas of high stress in need of relief. Comparison of ground-based information with a Landsat TM-derived land-cover map is used for accurate designation of agricultural regions in the study area. All dates before NASA MODIS availability (2000) are covered by NOAA Advanced Very High Resolution Radiometer (AVHRR) data. NASA is establishing criteria for comparing the MODIS data to the NOAA AVHRR NDVI data as well.

Only through careful integration of these complex and extensive data sources can NASA meet the critical challenge of lessening the destruction caused by drought and flood. Example exploratory questions one might ask that a data-mining tool provide data to answer include: "Given the seasonal forecast of temperature, humidity and precipitation for the next growing season and the past history of land surface vegetation coverage and greenness, what is the potential production by the end of the growing season?" or "How much is the change in the greenness (NDVI) in North America correlated to the El Niño as predicted by the seasonal forecast model?"

Possible Data Sources for climate variability and agriculture

MODIS (Terra and Aqua)
http://daac.gsfc.nasa.gov/MODIS/

TRMM
http://daac.gsfc.nasa.gov/hydrology/instrument_trmm.shtml

AVHRR
http://daac.gsfc.nasa.gov/guides/GSFC/guide/avhrr_dataset.gd.html

I.2.2.b Data preparation for medium range weather forecast

During the 1980s and 1990s, the ESE evolved the interdisciplinary field of Earth system science, deploying the Earth Observing System of sensors. Scientists use space-based observations, coupled with suborbital and in situ observations, to model the climate system. Global model and data assimilation systems, however, cannot ingest and utilize the large volume of satellite data relevant to a selected science question. Furthermore, it has been proven that not all data

contribute positively to the resulting model analysis and prediction impact. Model grids are typically much coarser than the data resolution (pixel size) of most remote sensing instruments. The data averaging within a grid filters the significant features. Typically, the data must be "thinned" to the model resolution and to filter out the non-contributing data, which detract from the desired model outcome.

Data thinning can be achieved by first identifying a desired feature and then subsetting the data around that target. For example, identify an atmospheric circulation pattern within a global data set and preserve those significant data values as input to a medium range weather forecast model. In this example, a proposer is asked to demonstrate a positive impact in forecast capabilities based on the improved data selection. Criteria for eliminating non-contributing data include sensor limitations (e.g. cloud-free data) and geographic limitations (e.g. exclude or only include the tropics). A typical data mining question for the weather forecast modeling community might be "How can sea surface winds, temperature, or topography be used effectively to improve medium range weather forecasts?" The outcome would be a data mining tool to enable forecast center modelers to describe desirable features or patterns to preserve in the remote sensing data inputs for the medium range weather forecast.

Possible Data Sources for the medium range weather forecast

Atmospheric Infrared Sounder:
http://www-airs.jpl.nasa.gov/

Ocean surface topography:
http://podaac.jpl.nasa.gov/cgi-bin/dcatalog/summary.pl?ost

Ocean vector winds:
http://podaac.jpl.nasa.gov/cgi-bin/dcatalog/summary.pl?ovw

Sea surface temperature:
http://podaac.jpl.nasa.gov/cgi-bin/dcatalog/summary.pl?sst

II. Proposal Technology Readiness Level (TRL) Requirement

For this mini-solicitation, the entry TRL shall be TRL 4 or greater. The proposal shall demonstrate that significant elements of the technology approach have been implemented and tested in full-scale problems or data sets (although not necessarily applied to the Earth science domain). The proposer shall identify the entry TRL, the planned exit TRL, and success criteria in their proposal. Past and ongoing work on the research activity should determine the entry TRL; the proposer shall substantiate the entry TRL in the proposal.

The proposal shall demonstrate at least one TRL advancement over the duration of the research; if the proposal duration for more than a year, one TRL advancement per year is desirable. TRL definitions can be found in Appendix F of the Master NRA document, NNG04ZY4000N.

III. Funding

The funding available for each mini-solicitation under this Master NRA will limit the number and magnitude of the proposals awarded. The ESTO expects that a total of 2 to 5 proposals will be awarded, with values in the approximate range of $50,000 to $300,000 per year per award.

NOTE: The NASA Grants and Cooperative Agreements Handbook, Section A, 1260.4 (b), specifies that awards to commercial firms via grants and cooperative agreements require 50% cost sharing when the commercial firm is expected to receive substantial compensating benefits for performance of the work.

Funds are not currently available for awards under this Master NRA. The Government's obligation to make award(s) is contingent upon the availability of appropriated funds from which payment can be made and the receipt of proposals that NASA determines are acceptable for award under this Master NRA. No additional funds above the initially specified award value will be available.

IV. Period of Performance

Awards will be made for the full period of performance, not to exceed 24 months. Grants and cooperative agreements are subject to annual review according to the criteria specified in the NASA Grant and Cooperative Agreement Handbook (14 CFR 1260). Proposals must define clear, measurable achievements for each year of performance.

V. Schedule

Proposals are due no later than 60 days after the official release date of this mini-solicitation. All proposals submitted in response to this mini-solicitation are due in accordance with the following schedule. Late proposals will not be considered. Proposals are to be submitted according to instructions in NNG04ZY4000N.

A complete proposal schedule is given below:

Proposals due	July 6, 2004, no later than 5:00 p.m. Eastern time
Peer Review	July 27, 2004
Announcement of Final Selections	August 6, 2004

Acronym List

AIST	Advanced Information Systems Technology
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
DAAC	Distributed Active Archive Center
ENSO	El Niño Southern Oscillation
EOS	Earth Observing System
ESE	Earth Science Enterprise
ESTO	Earth Science Technology Office
EVI	Enhanced Vegetation Index
GCM	Global Climate Model
GIS	Geographic information System
GSFC	Goddard Space Flight Center
LIDAR	Light Detection and Ranging
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aerospace and Aeronautics Administration
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanographic and Atmospheric Administration
NODC	National Oceanographic Data Center
NRA	NASA Research Announcement
OCTS	Ocean Color and Temperature Scanner
RCM	Regional Climate model
SEABASS	SeaWiFS Bio-optical Archive and Storage System
SeaWiFS	Sea-Viewing Wide Field-of-view Sensor
TRL	Technology Readiness Level
TRMM	Tropical Rainfall Measuring Mission