NASA SBIR 2007 Phase I Solicitation

S6 Information Technologies

Modeling and simulation are being used more pervasively and more effectively throughout NASA, for both engineering and science pursuits, than ever before. These are tools that allow high fidelity simulations of systems in environments that are difficult or impossible to create on Earth, allow removal of humans from experiments in dangerous situations, and provide visualizations of datasets that are extremely large and complicated. Examples of past simulation successes include simulations of entry conditions for man-rated space flight vehicles, visualizations of distant planet topography via simulated fly-over and three-dimensional visualizations of coupled ocean and weather systems. In many of these situations, assimilation of real data into a highly sophisticated physics model is needed. Also use NASA missions and other activities to inspire and motivate the nation's students and teachers, to engage and educate the public, and to advance the scientific and technological capabilities of the nation.

Subtopics

S6.01 Modeling, Simulation and Analysis Technologies

Lead Center: ARC

This subtopic solicits proposals for technologies and systems that allow spacecraft and ground systems to robustly perform complex tasks in dynamic environments with minimal human direction. Areas of interest include support of decision support systems, distributed sensor webs and component systems, and the creation of automation loops connecting scientific modeling and analysis to mission planning, data collection, processing and operations. NASA is moving from a stove-pipe observational architecture to one that permits data interoperability and dynamic coordination of observational assets to generate desired data products. Technology innovations include:

- Automation and autonomous systems that support high-level command abstraction;
- Efficient and effective techniques assessing gaps in data collection to assure complete coverage;
- Intelligent searches of distributed data archives, and data discovery through searches of heterogeneous data sets and architectures; and
• Automation of routine, labor intensive tasks to that either increase reliability or throughput of current process.

Specific areas of interest include the following:

• Search agents that support applications involving the use of NASA data using emerging interoperability such as Sensor Model Language;

• Methods that support the planning and scheduling of sensor webs in support of data product processing when given a set of high-level goals and constraints;

• Autonomous data collection including the coordination of space or airborne platforms while adhering to a set of data collection goals and resource constraints;

• System and subsystem health and maintenance, both space- and ground-based;

• Distributed decision making, using multiple agents, and/or mixed autonomous systems;

• Automatic software generation and processing algorithms; and

• Control of Field Programmable Gate-Arrays (FPGA) to provide real-time products.

S6.02 Technologies for Large-Scale Numerical Simulation

Lead Center: ARC
Participating Center(s): GSFC

NASA scientists and engineers are increasingly turning to large-scale numerical simulation on supercomputers to advance understanding of Earth and astrophysical systems, as well as to conduct high-fidelity engineering analyses. The goal of this subtopic is to make NASA’s supercomputing systems and associated resources easier to use, thereby broadening NASA’s supercomputing user base and increasing user productivity. Specific objectives are to:

• Reduce the learning curve for using supercomputing resources;

• Minimize total time-to-solution (i.e., time to discovery, understanding, or prediction);

• Increase the scale and complexity of computational analysis and data assimilation;

• Accelerate advancement of system models and designs.
The approach of this subtopic is to develop intuitive, high-level tools, interfaces, and environments for users, and to infuse them into NASA supercomputing operations. Successful technology development efforts under this subtopic would be considered for follow-on funding by, and infusion into either of the NASA high-end computing projects, including the High End Computing Columbia (HECC) project at Ames and the NASA Center for Computational Sciences (NCCS) at Goddard. Research should be conducted to demonstrate technical feasibility during Phase 1 and show a path toward a Phase 2 prototype demonstration. Open Source software and open standards are strongly preferred.

Specific areas of interest include:

### Application Development

With the increasing scale and complexity of supercomputers, users must often expend a tremendous effort to translate their physical system model or algorithm into a correct and efficient supercomputer application code. This subtopic element seeks intuitive, high-level application development environments, ideally leveraging high-level programming languages (e.g., parallel Matlab or IDL) to enable rapid supercomputer application development, even for novice users. This environment should dramatically simplify application development activities such as porting, parallelization, debugging, scaling, performance analysis, and optimization.

### Results V&V

A primary barrier to effective use of supercomputing by novices is understanding the accuracy of their computational results. Errors in the input data, domain definition, grids, algorithms, and application code can individually or in combination produce non-physical results that a user may not detect. This subtopic element seeks tools and environments to help users with verification and validation (V&V) of simulation results. This could be accomplished by enabling comparison of results from similar applications or with known accurate results, access to results analysis tools and domain experts, or access to error estimation tools and training.

### Data Analysis and Visualization

Supercomputing computations almost invariably result in tremendous amounts of data, measuring in the gigabytes or terabytes, and with many dimensions and other complexity aspects. This subtopic element seeks user-friendly tools and environments for analysis and visualization of large-scale, complex data sets typically resulting from supercomputing computations.
Ensemble Management

Conducting and fusing the results from an ensemble of related computations is an increasingly common use of supercomputers. However, ensemble computing and analysis introduces a new set of challenges for deriving full value from using supercomputing. This subtopic element seeks tools and environments for managing and automating ensemble supercomputing-based simulation, analysis, and discovery. Functions could include managing and automating the computations, model or design optimization, interactive computational steering, input and output data handling, data analysis, visualization, progress monitoring, and completion assurance.

Integrated Environments

The user interface to a supercomputer is typically a command line or text window, where users may struggle to locate or develop applications, understand the job queue structure, develop scripts to submit jobs to the queue, manage input and output files, archive data, monitor resource allocations, and many other essential supercomputing tasks. This subtopic element seeks more intuitive, intelligent, and integrated interfaces to supercomputing resources. This integrated environment could include access to user training (e.g., tutorials, case studies, and experts), application development tools, standard (e.g., production, commercial, and Open Source) supercomputing applications, results V&V tools, computing and storage resources, ensemble management tools, workflow management, data analysis and visualization tools, and remote collaboration.

S6.03 On-Board Data Processing and Control

Lead Center: ARC

Participating Center(s): GSFC

Technology advances allow scientists to build devices that often collect more data than can be cost effectively transmitted or summarized within mission time constraints. NASA is developing sensor web capabilities which can require these data be analyzed for rapid decision making, either autonomously or with human in the loop controller. This subtopic enables sensor web capabilities and increases mission data return by developing on-board methods that can operate with very limited resources to increase the efficiency and scientific return of existing and future sensors. Approaches range from losses less data compression prior to transmission to some degree of "data understanding" that enables data management and prioritization based on potential science content. These software capabilities will enable sensor webs that operate semi-autonomously and are capable of reacting to what is being sensed and triggering notifications or additional actions. Algorithms can be embedded into an instrument or device or algorithms can target on-board computer resources for data management and/or transmission as part of the post collection data flow.

The selection of on-board methods to increase scientific return is highly dependent on mission objectives. Successful candidate technologies will need to demonstrate suitability to the general requirements of the proposed use scenario as they pertain to different instrument (or device) types. Generally, scientists do not want to throw away data given that significant discoveries have been made reinterpretation archived data. Methods that reduce information content such as lossy compression are often not desirable unless significant, new capabilities are
enabled by this tradeoff. Examples exist where instruments are turned off and on and instances when sensor or camera data is saved and transmitted only when features are detected by on-board software. These instances occur when transmission costs, relative to available resources, are high. E.g., a Mars Exploration Rover was reprogrammed to detect and transmit camera images containing dust devils.

Algorithms can be designed to run on general purpose computing resources or specialized i.e., field programmable gate arrays (FPGA). Novel approaches that can leverage specialized, space qualified computing resources such as FPGAs that return order of magnitude reduction in data volume or screening capabilities are desirable. There is a trade-off between sensor volume and complexity against distance and degree of on-board autonomy needed for mission success so performance metrics are relative to the science mission scenario. Example sensor types include data intensive instruments such as hyperspectral, RADAR, and LIDAR but can include any sensor technology that is shown relevant to the board scope of science within the NASA science mission directorate.

For instance, aggressive metrics for compression and data volume are in Earth science the Decadal survey has the following requirements on data compression:

<table>
<thead>
<tr>
<th>RADAR Missions</th>
<th>SMAP (RADAR)</th>
<th>DESDynI (RADAR)</th>
<th>SWOT (RADAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBP Input data rate (MHz)</td>
<td>32</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>Processor Throughput (GFLOPS)</td>
<td>7</td>
<td>20</td>
<td>90</td>
</tr>
<tr>
<td>Data Compression Ratio</td>
<td>80:1</td>
<td>10:1</td>
<td>90:1</td>
</tr>
</tbody>
</table>

Where raw data sample spacing is 0.75 m x 1.5 m (16 bits per sample), and the output data sample spacing is 10 m x 10 m (16 bits per sample).

For Hyperspectral imaging instruments, here is an exemplar requirement on data compression on-board feature detection.

<table>
<thead>
<tr>
<th>Data Rate:</th>
<th>660 gigabits per orbit, 220 megabits per second</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Compression Ratio:</td>
<td>&gt; 3.0</td>
</tr>
<tr>
<td>On-board detection capability:</td>
<td>A quick look at the data for presence of cloud cover.</td>
</tr>
</tbody>
</table>

**S6.04 Data Analyzing and Processing Algorithms**

**Lead Center: GSFC**

This subtopic seeks technical innovation and unique approaches for the processing and analysis of data from NASA’s space and Earth science missions. Analysis of NASA science data is used to understand dynamic systems such as the sun, oceans, and Earth’s climate as well as to look back in time to explore the origins of the universe.
Algorithms are used to consider data over time, at various energy ranges, and at different points in space. Complex algorithms and intensive data processing are needed to understand and make use of this data. What novel discoveries can be made with existing NASA data? What applications would benefit from the combination of NASA data with additional information and processing?

NASA seeks to exploit spatial tools in order to increase the utility of scientific research data, models, simulations, and visualizations. Of particular interest are innovative computational methods to dramatically increase algorithm efficiency and thus performance. Interpolation, clustering, and registration algorithms are examples of the type of algorithms of interest in this area, as well as real-time visualization and simulation algorithms. Tools to improve predictive capabilities, to optimize data collection by identifying gaps in real-time, and to derive information through synthesis of data from multiple sources are needed. The ultimate goal is to increase the value of data collected in terms of scientific discovery and application. Data analysis and processing must relate to advancement of NASA's scientific objectives.

We are soliciting proposals for software tools which access, fuse, process, and analyze image and vector data for the purpose of analyzing NASA’s space and Earth science mission data. Tools can be plug-ins or enhancements to existing software or on-line services. Tools and products might be used for broad public dissemination or for communicating within a narrower scientific community. Tools can be new stand-alone applications or web services, provided that they are compatible with most widely-used computer platforms and exchange information effectively (via standard protocols and file formats) with existing, popular applications. The Phase 1 contract should demonstrate the feasibility of the approach. The Phase 2 contract should provide prototype software that can be demonstrated at the company and a prime contractor or NASA. It is desirable to have the development lead to software that is commercialized or infused into NASA program use.

To promote interoperability, tools shall use industry standard protocols, formats, and APIs, including compliance with the ISO, FDGC, and OGC standards as appropriate. For example a tool may manipulate XML of various types, such as GML, SensorML, KML; or use standard services, such as WSDL and UDDI. Applications may subset, filter, merge, and reformat existing spatial data; provide links to attribute data; or visualize results. Combining NASA research data with popular geospatial services is encouraged.

S6.05 Data Management - Storage, Mining and Visualization

Lead Center: GSFC

This subtopic focuses on supporting science analysis through innovative approaches to managing and visualizing collections of science data which are extremely large, complicated, and are highly distributed in a networked environment that encompasses large geographic areas. There are specific areas for which proposals are being
sought:

3D Virtual Reality Environments

- 3D virtual reality environments for scientific data visualization that make use of novel 3D presentation techniques that minimize or eliminate the need for special user devices like goggles or helmets;
- Software tools that will enable users to "fly" through the data space to locate specific areas of interest.

Distributed Scientific Collaboration

- Tools that enable high bandwidth scientific collaboration in a wide area distributed environment;
- Novel tools for data viewing, real-time data browse, and general purpose rendering of multivariate geospatial scientific data sets that use geo-rectification, data overlays, data reduction, and data encoding across widely differing data types and formats.

Distributed Data Management and Access

- Metadata catalog environments to locate very large and diverse science data sets that are distributed over large geographic areas;
- Dynamically configurable high speed access to data distributed and shared over wide area high speed network environments;
- Object based storage systems, file systems, and data management systems that promote the long term preservation of data in a distributed online (i.e., disk based) storage environment, and provide for recovery from system and user errors.

Research should be conducted to demonstrate technical feasibility during Phase 1 and show a path toward a Phase 2 hardware/software demonstration, and when possible, deliver a demonstration unit for functional and environmental testing at the completion of the Phase 2 contract.
This subtopic seeks technical innovation and unique approaches to exploit spatial tools in order to increase the use of NASA research data, models, simulations, and visualizations. The goal is to facilitate NASA’s Science and Exploration Missions, and outreach to the interested public. These tools will be used by the NASA Applied Sciences Program managed by the Applied Research and Technology Project Office at Stennis Space Center. The tools should be easy to use by non-specialists, from scientists and policy makers to the general public. Tools and services will be prototyped for accessing and fusing (or mashing) image and vector data with popular Web-based or stand-alone applications. Tools can be plug-ins or enhancements to existing software or on-line services. Tools and the products might be used for broad public dissemination or for communicating within a narrower scientific community.

For example, an authoring tool may help a non-GIS expert to map a National Weather Service modeled hurricane path over a background of NASA MODIS sea surface temperatures, in turn draped on a visualization of the globe served by GoogleEarth.

To promote interoperability, tools shall use industry standard protocols, formats, and APIs. For example a tool may manipulate XML of various types, such as GML, SensorML, KML; or use standard services, such as WSDL and UDDI. Applications may subset, filter, merge, and reformat existing spatial data; provide links to attribute data; or visualize results. Combining NASA research data with popular geospatial services is encouraged. Examples of popular applications and services currently include:

- Imagery servers: e.g., NASA DAACs, OGA servers (USGS, NOAA, DOI), Microsoft Terraserver, Google Maps;
- Mapping platforms: e.g., Google Earth, NASA WorldWind;
- Map servers: e.g., Census Bureau, EPA Maps, Google Maps, MapQuest, Yahoo Maps.