PROJECT OVERVIEW

The overall goal of the National Aeronautic and Space Administration’s (NASA) initiative to create Rapid Prototyping Capability (RPC) is to speed the evaluation of potential uses of NASA research products and technologies to improve future operational systems by reducing the time to access, configure, and assess the effectiveness of NASA products and technologies. The developed RPC infrastructure will accomplish this goal and contribute to NASA’s Strategic Objective to advance scientific knowledge of the Earth system through space-based observation, assimilation of new observations, and development and deployment of enabling technologies, systems and capabilities including those with potential to improve future operational systems.

The infrastructure to support the RPC is thus expected to provide the capability to rapidly evaluate innovative methods of linking science observations. To this end, the RPC should provide the capability to integrate the tools needed to evaluate the use of a wide variety of current and future NASA sensors and research results, model outputs, and knowledge, collectively referred to as “resources”. It is assumed that the resources are geographically distributed and thus the RPC will support location transparency of the resources.

SYSTEM REQUIREMENTS

The RPC infrastructure will support at least two major categories of experiments (and subsequent analysis): comparing results of a particular model as fed with data coming from different sources, and
comparing different models using the data coming from the same source.

In spite of being conceptually simple, two use cases in fact entail a significant technical challenge. The barriers currently faced by the researchers include inadequate data access mechanisms, lack of simulated data approximating feeds from sensors to be deployed by future NASA missions, plethora of data formats and metadata systems, complex multi-step data pre-processing, and rigorous statistical analysis of results (comparisons between results obtained using different models and/or data).

Enabling RPC experiments, in this context, means thus a radical simplification of access to both actual and simulated data, as well as tools for data pre- and post-processing. The tools must be interoperable, allowing the user to create computational workflows with the data seamlessly transferred as needed, including third-party transfers to high-performance computing platforms. In addition, the provenance of the data must be preserved in order to document results of different what-if scenarios and to enable collaboration and data sharing between users.

The development of RPC system does not involve developing the tools for data processing. These tools are expected to be provided by the researchers performing experiments, projects focused on the tool development, and the community at large. Indeed, many tools for handling Earth science data are available from different sources, including NASA, USGS, NOAA, UCAR/Unidata, and numerous universities. Instead, the RPC system is expected to be an integration platform supporting adding ("plugging in") tools as needed.

The essence of the RPC process is to provide an evaluation of feasibility of transferring research capabilities into routine operations for societal benefits. The evaluation should result is a recommendation for the NASA administrators to pursue or abandon the topic under investigation. Since making an evaluation requires a narrow expertise in a given field (e.g., invasive species, crop predictions, fire prediction, etc.), the results presented by a particular evaluation endeavor needs to be peer-reviewed. The proposed approach is to provide means for publishing the results electronically – that is, providing the community access not only to the final reports and publication but also to the data used and/or produced during the analysis, as well as providing access to tools used to derive the conclusions of the evaluation.

IMPLEMENTATION

The development of an RPC system satisfying all the requirements described above is an immense task. Consequently, one of the most important design decisions was to prioritize the system features and selecting the sequence of actions that would lead towards the implementation of the full functionality:

**Phase I:** Interactive web site for describing the experiments and gathering the feedback form community

**Phase II:** RPC data server acting as a cache for experimental data (by extending the functionality of the Unidata’s THREDDS server).

**Phase III:** Online tools for data processing ("transformations") such as HEG, MRT, TSPT, and ART for format translations, reprojections, subsetting and stitching, generation of simulated VIIRS data sets, creating time series, as well as performance metrics tools such as visualizations and statistic analysis tools.

**Phase IV:** Support for batch processing (asynchronous bulk processing on remote high-performance platform)

**Phase V:** Creation of a Virtual Organization for sharing data, codes, and computational resources.

The functionality of the RPC system naturally splits into several independent modules such as interactive Web site, data server, tool’s interfaces, or monitoring service. Each such module is implemented as an independent portlet, aggregated by the RPC Portal. The system implements the Service-Oriented Architecture realized by open-source Service Bus (allowing easy integration of new components) and Globus Grid Computing Toolkit (enabling access to remote resources).

*This work was supported by a contract from NASA/SSC under the Earth Sciences Program.*