

ESIP Products and Services Committee 2011 Budget Request

The Products and Services Committee is requesting \$32,000 for FY 2011 to improve and expand its internal testbed that evaluates prototype standards, services, protocols, and best practices. These capabilities help users to *discover, use, and interpret Federation products and services*. The testbed serves as a forum for innovative collaboration across all sectors of the Federation by enabling Federation committees and clusters to develop, test and improve availability and accessibility of member products and services. In addition to newly proposed test activities, this year we also propose to develop a new content management system-based **ESIP Testbed portal** to host the Testbed activities as well as share comments about each Testbed topic, including suggested best practices. The Testbed is intended to offer a unified environment to gain consensus on community best practices and approaches to utilizing technological advances that are of interest to the ESIP Federation community.

Each of this year's proposed Test activities is being **hosted/recommended by other ESIP Federation clusters or committees**. This Federation-wide involvement will hopefully serve to boost participation and interest in the Testbed activities beyond the Product and Services Committee. The improved level of participation will hopefully make the Testbed more viable and successful.

To provide oversight and long-term sustainability of the Testbed, the Products and Services Committee will convene a **Testbed Configuration Board (TCB)** composed of members from at least five (5) interested ESIP Federation committees and clusters, as well as representation from the ESIP Federation front office. The TCB, in the future, will review and approve request for new Testbed activities as well as provide guidance, oversight and direction of the Testbed.

Appendix A lists the proposed testbed tasks. The testbed will require computer and web programming support (Contract Services) at the 0.5 FTE level to establish the infrastructure for all tasks. All task-related expenses are for **Contract Services**.

To determine who will carry out each funded task, a request will be sent out to the ESIP-all mail list for brief proposals. Any member of the ESIP community may submit a proposal or be part of a proposal team. The proposals will be evaluated by a subcommittee within the Products and Services Committee. Winning teams will be required to provide written **monthly progress reports**, participate on Committee telecons and to present a poster or other **report at a Federation Meeting**. The Committee will provide **semi-annual written reports** of accomplishments to the Executive Committee and brief oral reports at Federation Meetings. The reports will highlight the value added to Federation member products and services accruing from the Testbed.

These tasks will build upon those carried out in FY2009 that established the testbed. As part of that activity, the testbed enabled: i) a searchable member directory by skills; ii) a

testing of unique object identifier schemes to assign a permanent name to Federation member datasets; iii) a community ontology of data services and data types; and iv) a clone of the Air Quality portal for other application areas. We did not request any funds for FY2010.

The FY2011 request aligns with the following Federation Strategic Plan elements:

- Demonstrate use through community-vetted demos, pilots and applications.
- Provide mechanisms for community review of data, products, applications and other resources.
- Develop and share alternative approaches to sustaining Earth science data and information networks
- Promote use of technical standards and best practices for data management, stewardship and application development.
- Recognize and promote best practices for providing feedback to sponsors.

Appendix A: Proposed Testbed Tasks

1. Deploy a Testbed Portal

Currently access to testbed activities is primarily provided through links on the Products and Services Committee section of the ESIP Federation Wiki. While this approach works fine if users realize the Testbed is sponsored by the Products and Services Committee, it is otherwise somewhat hidden and difficult to discover from outside of the wiki environment. We propose to pursue **two objectives** with a new Testbed environment; (1) In an effort to make the ESIP Testbed more visible and accessible to ESIP members and the general public, a CMS-based **portal** will be deployed that will host the various funded activities and be available to support other activities as deemed appropriate by the Testbed Configuration Board. Access controls on the portal will be configured to provide controlled access for developers and testers while at the same time providing public access for users and other participants. (2) Additionally, we propose to implement a **new flexible operating environment** for the Testbed that will allow for more customized configuration for each task (if necessary) through the use of *virtualization services*, as well as making use of *cloud-based resources* as available. Some tasks will have requirements for system dependencies that may collide with other tasks, so the availability of virtualized systems will allow more specialized configurations for tasks with those requirements. A common operating area will be available to host the less complicated configuration of other tasks as necessary. The Testbed Configuration Board will manage long-term maintenance and configuration of the Testbed portal.

ESIP Community Benefits: Better accessibility of the ESIP Federation Testbed.

Cost: \$5K to develop, test and deploy the Testbed portal.

2. Discovery Services and Clients: Interoperability Testing, Advertisement, and Assessment of Data to Service Broker

The Discovery cluster provides a medium for Federation members to coordinate on development, deployment, and creation of interoperable specifications for **Discovery services such as OpenSearch, DataCasting, and ServiceCasting**. As such, a common testbed to deploy services and/or clients would enable the Federation members to test the interoperability of Discovery services and clients. More over, having a common testbed where Federation members from various data centers can access would provide a common platform that can be used to advertise their services. Currently it is difficult to know where services exist, what they do, and how to access them. Several machine-processable mechanisms for advertising services (such as service casting) have been evaluated by the NASA Tech Infusion Working Group. The testbed will provide an environment to demonstrate how Federation-wide member services can be advertised, described, and chained together using existing technologies. From this, new capabilities can also be explored such as enabling an

association of data to services where for each data collection or granule, compatible services provided by Federation members can be shown.

ESIP Community Benefits: Improved interoperability of ESIP Discovery services and clients leading to broader adoption. Increased awareness and usage of Discovery services. Improved brokering of data collection and granules to compatible services that can use them.

Cost: \$8K to (1) setup the testbed for allowing service and client implementations to be deployed onto the testbed, (2) setup a service cast of Federation member's services offered, and (3) explore concept of data to service mapping.

3. Data and Information Quality

Many groups within ESIP have longstanding interests in providing quality measures for Federation products and services. Data quality has many dimensions (instrument accuracy, algorithm effectiveness, etc.) and is highly dependent upon the intended applications. For example, real-time applications have relatively low **accuracy** requirements, but **service quality** (% of time operational) is crucial. We will establish criteria for various dimensions and levels of services of data and service quality. Quality annotations will be elicited from both ESIP providers and the greater user community who can "rate" Federation offerings (1 to 5 stars) and provide supporting comments.

ESIP Community Benefits: Better branding and usability of Federation products and services

Cost: \$6K to setup a classification/annotation scheme for data and service quality according to standard quality measures and dimensions.

4. Develop Re-usable Authoring Tools for FGDC/ISO-compliant Metadata

The ESIP Information Technology and Interoperability (IT&I) Committee proposes the development of a suite of metadata authoring tools that will be reusable across Earth science data repositories. The tools will be developed as a Drupal profile (GeoDrupal), that can be used by data producers to create/author metadata information for geospatial data sets, either through an easy-to-use, interactive, web-based interface, or by uploading XML definitions. The resulting metadata will be FGDC/ISO compliant. Additionally, this task will investigate the development of metadata translation tools to assist in the conversion of existing metadata to other formats. Several Earth science data centers utilize an internal metadata format that is often useful to have converted to additional standards (i.e. GCMD DIF & SERF, ECHO, FGDC, etc.) The software will be published and shared at Drupal.org on behalf of the ESIP Federation, making it freely available for deployment at multiple sites. The modules necessary for this functionality will be incrementally designed and implemented under the direction of the IT&I committee.

ESIP Community Benefits: Provide re-usable metadata authoring tools.

Cost: \$6K to develop, test and deploy within the Testbed portal.

5. Data Stewardship Testbed Activity

The Preservation and Stewardship Cluster and the NASA Technology Infusion Working Group have been considering permanent identifier schemes for data products http://wiki.esipfed.org/index.php/Preservation_and_Stewardship. These identifiers can serve as references in journal articles as well as inventory nodes in data archives and must include representations for versions of the entity being identified. Many identifier options have been proposed for different kinds of data, but the best choices for Earth science data require careful examination. For example, two datasets may differ only in format, byte order, data type, access method, etc., creating distinctions between them that may not be addressed adequately by identifier schemes used for typical "published" items such as books and journals. Previous Testbed activity in this area included a recommendation on identifier schemes to use for Earth Science data, but did not address the implementation issues that arise with the identifier schemes considered. The proposed Task for this work is to examine several different kinds of datasets, assign identifiers from up to nine identifier schemes considered in the previously mentioned paper, evaluate and compare the implementation implications and other practical considerations associated with the use of each identifier scheme applied, and develop recommendations. Practical considerations may include the need to integrate with other metadata schemes such as ISO, and application to data citation formats and practices.

ESIP Community Benefits: Permanent, unique names for Federation data products and recommendations for practice based on testbed experience.

Cost: \$7K to develop, test and deploy Testbed interfaces within the portal for this activity.