Earth Science Mining Web Services

**Abstract** To allow scientists further capabilities in the area of data mining and web services, the Goddard Earth Sciences Data and Information Services Center (GES DISC) and researchers at the University of Alabama in Huntsville (UAH) have developed a system to mine data at the source without the need of network transfers. The system has been constructed by linking together several existing technologies: the Simple Scalable Script-based Science Processor for Measurements (S4PM), a processing engine at the GES DISC; the Algorithm Development and Mining (ADaM) system, a data mining toolkit from UAH that can be configured in a variety of ways to create customized mining processes; ActiveBPEL, a workflow execution engine based on BPEL (Business Process Execution Language); XBaya, a graphical workflow composer; and the EOS Clearinghouse (ECHO).

**Web Services enables the infusion of diverse technologies**

XBaya // mining workflow composer
- User authors workflow and deploys to ActiveBPEL engine.

ActiveBPEL // workflow orchestration engine
- Exposes a URI pointing to the WSDL for that workflow
- Workflow URL is sent to the GES DISC Data Mining Services via SOAP
- Discovery service for data files
- Utilizes ActiveBPEL to execute ADaM via web service according to workflow
- Stages output to an FTP area for pickup by external user.

Simple, Scalable, Script-Based Science Processor for Measurements (S4PM):
- Processing engine searches ECHO for data, and caches locally
- Invokes ActiveBPEL to execute ADaM via web service according to workflow
- Stages output to an FTP area for pickup by external user.

**Conclusion**

Earth Science Mining Web Services, created from an infusion of well-known technologies, have shown promising results to the data mining/scientific community. With an abundance of algorithms available, users can create and execute their data mining workflows without any data transfer. In turn this gives user control over the data they want to process at the server's source. The next phase will be the Smart Assistant for Earth Science Data Mining (SAM). SAM will provide data type mining ontologies to aid in workflow composition, expansion of existing workflow composer tool and deployment of existing mining services in additional environments.