CTS 2009 DEMO SESSIONS

DEMO III

Collaborative Workflows in Earth Science Data Mining

Chris Lynnes  
NASA Goddard Space Flight Center  
Greenbelt, Maryland, USA

Rahul Ramachandran, Helen Conover, Sunil Movva  
ITSC/University of Alabama - Huntsville  
Huntsville, Alabama, USA

Brian Wilson  
NASA Jet Propulsion Lab  
Pasadena, California, USA

ABSTRACT

A small but growing number of scientists and researchers are beginning to harness Web 2.0 technologies as a transformative way of doing science. Since communication is at the heart of science, these technologies provide researchers easy mechanisms to share ideas, data and algorithms. These technologies complement formal means of sharing knowledge via conferences and published papers, where it is impossible to share all the research details, and where negative results are rarely included. Meanwhile, science software developers are converting data processing, analysis, mining and visualization algorithms into publicly available web services, allowing researchers access to large suites of algorithms for data processing and science analysis. This model of chaining services to create analysis workflows provides the research community unprecedented opportunity to collaborate, sharing their workflows, reproducing and analyzing research results, and leveraging colleagues’ expertise to expedite the process of scientific knowledge discovery. In many cases, the output of one workflow can be an input to others, leading to chained workflows with components shared by two or more researchers.

A crucial component to foster this unprecedented cooperation within the research community is a reusable, extensible and customizable environment for building collaborative “open science” portals to manage these shared analysis workflows. Current collaborative portals are often one-time development efforts for specific science domains that cannot be easily extended beyond their initial features or reused by other science domains. As part of a current NASA project, we are developing “Talkoot” (Finnish analogue to a barn raising): a customizable “software appliance” to build
collaborative portals for Earth Science services and analysis workflows. Talkoot will allow researchers (not just information technologists) to build collaborative sites around service workflows within a few hours. Talkoot leverages Drupal, an open architecture Content Management System, as the framework for an online collaborative portal. Drupal also has a number of specialized modules developed by its user community to provide additional features. Talkoot adds Earth-science-specific modules to provide data searching, processing, analysis and mining capabilities. We will demonstrate the current prototype of Talkoot, which allows users to: create research experiments; select data sets; create data processing/analysis/mining workflows; execute these workflows on remote compute resources; and share both the results and the workflows with other users. Additional Talkoot features that are being developed or are planned will also be presented.

PRESENTER
Chris Lynnes, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

PRESENTER BIOGRAPHY
Dr. Lynnes is the Chief Systems Engineer at the Goddard Earth Sciences Data and Information Services Center (GES DISC) at NASA’s Goddard Space Flight Center. He holds a Ph. D. in Geophysics (Seismology) from the University of Michigan. Dr. Lynnes has 20 years of experience in designing and implementing large-scale Earth science data and information systems for seismology and remote sensing.