

## Talkoot: Discover, Tag, Share, and Reuse Collaborative Science

Rahul Ramachandran  
University of Alabama in Huntsville  
*rramachandran@itsc.uah.edu*

On the emerging “Social Web,” millions of people offer their knowledge online in a collective knowledge system comprising an active community of motivated members posting problems and solutions in blogs, forums, mailing lists, collaborative portals and other Web 2.0 technologies. A small but growing number of scientists and researchers are beginning to harness these 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 critique, suggest, and 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.

At the same time, science software developers have embraced Service Oriented Architectures (SOA) and Systems of Systems such as GEOSS. Data processing, analysis, mining and visualization algorithms are being converted 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 with one another, 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 needed for this unprecedented level of cooperation within the research community is a reusable, extensible and customizable environment for building collaborative “open science” portals for managing these shared analysis workflows. Current collaborative portals (e.g., MyExperiment.org) have been one-time development efforts for specific science domains that cannot be easily extended beyond their initial features or reused by other science domains.

This paper describes the design and the development of a customizable “software appliance” to build collaborative portals for Earth Science services and analysis workflows which will greatly facilitate community collaboration in science analysis. The critical requirement of this software appliance is that researchers (not just information technologists) be able to build collaborative sites around service workflows (e.g., see Fig. 1) within a few hours. We envision online communities coming together, much like Finnish “talkoot,” to build a shared research space. Our Talkoot software appliance will be applicable to many different science domains, mission teams, research projects and organizations.

The Internet has enabled a new phase in collaboration among science researchers, such as the use of wikis for community research. However, wikis are limited to text, static

images and hyperlinks, providing little support for collaborative data analysis. We are developing Talkoot as a novel open science environment that integrates many Web 2.0 technologies into a turnkey software appliance for constructing collaborative portals, representing a major improvement in science collaboration tools. More importantly, with the ability to share and execute analysis workflows, Talkoot portals can be used to do collaborative science in addition to communicate ideas and results, foster the development of collaborative online communities to manage science workflows, and create a growing world of directly reproducible science to support exploratory science, informal group review, and even formal peer review.

Features of Talkoot include:

- **Visually authored, executable Web and Grid workflows** that preserve provenance, serve as templates for reuse and enhancement, and “grow” the world of reproducible science.
- **Social tagging** of all content for easy discovery and automated recommendations.
- **Intelligent, semi-automated workflow composition** that grows smarter based on a user’s social tagging and construction of simple taxonomies for their services and datasets.
- **Group-shared or public “science notebooks”** (webs of pages) that support all phases of science from exploration to final publication, and that can be group-edited, annotated by collaborators, commented upon by interested users, or even peer reviewed.
- **Dynamic science content** that can be modified at the push of a button, or triggered by an event, because each plot or figure is reproducible by the workflow that created it.
- **Continuous advertisement of the growing portal content** using RSS/Atom feeds: datasets by “datacasts”, web services and workflows by “service casts”, important geophysical events with linked datasets by “event casts”, and topics and groups by “interest casts”.

This paper will describe in detail the Talkoot vision of an open science collaborative environment, its design and implementation details.