

# **The Geoscience Paper of the Future Initiative: Training Scientists in Best Practices of Software Sharing**

Yolanda Gil (USC), Chris Duffy (PSU), Chris Mattmann (JPL), Erin Robinson (ESIP), Cedric David (JPL), Ibrahim Demir (UI), Bakinam Essawy (UV), Robinson W. Fulweiler (BU), Jon Goodall (UV), Leif Karlstrom (UO), Kyo Lee (JPL), Heath Mills (UH), Ji-Hyun Oh (USC), Suzanne Pierce (UT), Allen Pope (CU), Mimi Tzeng (DISL), Karan Venayagamoorthy (CSU), Sandra Villamizar (UC), Xuan Yu (UD)

*Contact email: gil@isi.edu*

## **Motivation: Training Scientists in Best Practices of Software Sharing**

As part of the EarthCube OntoSoft project, we set out to develop a software registry for geosciences with associated training materials for scientists to describe, publish, and disseminate their code so it can be reused by others. We formed an Early Career Advisory Committee from diverse areas of geosciences to provide the project with guidance on scientist requirements. Our focus was early career researchers who develop most of the software and at the same time can be an engine of change in their respective fields.

Very early on in the project we found that the Committee members were more interested in learning about best practices for software sharing than learning to improve their coding skills. In addition, we found that despite their interest it was hard for them in practice to make the time to take the training as software issues were not a high priority relative to everything else. What was needed is to embed the training into their work.

## **The Geoscience Paper of the Future: On the Job Training**

The Geoscience Papers of the Future (GPF) Initiative aims to train scientists in best practices of software and data sharing as they write a paper about their work. The Initiative has two major components. First, a Special Section of the AGU Earth and Space Science journal that will highlight GPFs. We have partnered with the largest scientific society in geosciences to create a special issue of their new cross-disciplinary journal to encourage submissions of GPFs. Submissions opened July 1, 2015 and close January 1, 2016. Second, a training program that offers training sessions for geoscientists to learn best practices in software and data sharing, provenance documentation, and scholarly publication. These training sessions will be offered in different modalities throughout 2015 at scientific meetings and selected institutions. After the initial training session given to OntoSoft Early Career Advisory Committee participants, the first batch of GPFs is being prepared to serve as exemplars for future authors.

A GPF satisfies the following requirements: 1) *Data requirements*: Making data available in a public repository, including documentation (metadata), a clear license specifying conditions of use, and citable using a unique and persistent identifier; 2) *Software requirements*: Making software available in a public repository, with documentation, a license for reuse, and a unique and citable using a persistent identifier; 3) *Provenance requirements*: Documenting the provenance of results by explicitly describing the series of computations and their outcome in a workflow sketch, a formal workflow, or a provenance record, possibly in a shared repository and with a unique and persistent identifier. Geoscientists are trained in best practices for these topics as they work on writing a GPF.