

Position Statement for the 2015 NSF Workshop on Intelligent Systems for Geosciences, March 2015

Yolanda Gil

Information Sciences Institute and Computer Science Department
University of Southern California
gil@isi.edu

My research is in intelligent systems, in particular intelligent user interfaces, Semantic Web, scientific workflows, metadata and provenance.

Prior to the workshop, we were asked to summarize 3-4 big ideas in your field in the last few years that you think may have an impact in geosciences research. In my view, there has not been a lot of work in AI that has impacted geosciences – and this is part of the motivation to hold this workshop. But for other sciences, they are:

- *Semantic Web technologies*: The adoption of Semantic Web concepts in the geosciences has been very limited, particularly when compared with the broad use of these technologies in biomedical research. However, there are massive amounts of relevant data already available in a form that is known as Linked Open Data (<http://www.w3.org/standards/semanticweb/data>), which uses Semantic Web languages to publish ontologies and data as open Web objects. These include geospatial data, sensor data, and other potentially relevant data for geosciences. A new joint W3C/OGC effort for spatial data on the Web will boost this work (<http://www.w3.org/2015/spatial/>).
- *Workflow systems*: Capturing complex data analysis processes as workflows facilitates reuse, scalable execution, and reproducibility. Workflow systems are used by hundreds of biomedical researchers, and thousands more in the biomedical industry. In geosciences, adoption has been limited but it is growing as workflow systems are becoming popular open source projects (e.g., <https://www.apache.org/>).

We were also asked to highlight 2-3 important research trends in your area that can be relevant to the workshop goals. These are the trends that I see:

- *Intelligent user interfaces for the Semantic Web*: Just like other complex infrastructure (eg cloud computing, web services), semantic technologies are easier to adopt with user interfaces that are accessible to scientists. My current work is extending semantic wikis to allow scientists to describe datasets, software, shared tasks, and field samples and export them as open Semantic Web Objects that can be interlinked and used in other systems (<http://www.socialknowledgecollection.org>). I believe that in coming years we will see a broader use of such interfaces to populate a massive Semantic Web for geosciences.
- *Intelligent workflow systems*: When workflows are augmented with expert-grade rules to select and customize analyses for any given data, automation and validation become possible. The pace of research could be significantly accelerated by these intelligent workflow systems running on data repositories.
- *Provenance*: Scientists always capture provenance, albeit mostly on informal notes in unstructured text. A major enabler to capturing provenance is the recent W3C PROV standard (<http://www.w3.org/2011/prov/>), which is now being integrated with OGC and ISO standards. As these standards are incorporated into the tools that geoscientists use, they will facilitate the discovery and reuse of data.