Geophysical Research Abstracts
Vol. 13, EGU2011-4928-3, 2011
EGU General Assembly 2011

Experiences Developing a User-centric Presentation of Provenance for a Web-based Science Data Analysis Tool

Stephan Zednik (1), Gregory Leptoukh (2), Peter Fox (1), Chris Lynnes (2), and Jianfu Pan (3)

(1) Tetherless World Constellation, Rensselaer Polytechnic Inst., Troy, NY, United States, (2) NASA Goddard Space Flight Center, Greenbelt, MD, United States, (3) NASA Goddard Space Flight Center/Adnet Systems, Inc., MD, United States

Web-based science analysis and processing tools allow users to access, analyze, and generate visualizations of data while alleviating users from having to directly manage complex data processing operations. These tools provide value by streamlining the data analysis process, but usually shield users from details of the data processing steps, algorithm assumptions, caveats, etc. Correct interpretation of the final analysis requires user understanding of how data has been generated and processed and what potential biases, anomalies, or errors may have been introduced. By providing services that leverage data lineage provenance and domain-expertise, expert systems can be built to aid the user in understanding data sources, processing, and the suitability for use of products generated by the tools.

We describe our experiences developing a semantic, provenance-aware, expert-knowledge advisory system applied to an existing web-based Earth science data analysis tool (e.g. Giovanni from NASA/GSFC). First, we introduce our use cases for provenance and science metadata capture, reasoning, and visualization to the end user. Then we introduce our integrated semantic data model, which is comprised of provenance, data processing, and science domain ontologies. We describe how we developed an initial set of expert rules, to reason over our data model and recognize conditions in the processing provenance that could lead to anomalies or errors in the processing results. Finally we will show how knowledge from the semantic data model and inferences of the advisory expert ruleset is presented to the user to assist in user understanding of the suitability of products generated by the analysis tool.