Geophysical Research Abstracts,
Vol. 11, EGU2009-13208, 2009
EGU General Assembly 2009

Semantic Provenance Management for Large Scale Scientific Datasets

P. Fox and D. L. McGuinness

Rensselaer Polytechnic Institute, World Tetherless Constellation, Troy, United States (pfox@cs.rpi.edu, +15182764464)

Our work to date on semantic provenance in real-world data production pipelines has lead to an emerging understanding of what the key annotations of provenance collection requirements are for a broad varity of end users. Having documented several of the image data pipelines for solar physics instruments at the Mauna Loa Solar Observatory as well as almost 20 use cases from the instrument scientist, observers, data analysts and managers, and end-user scientists, we have developed an initial infrastructure that can be searched, queried or browsed by these users for purposes of specific intereest to them. For example,

Our motivation behind a multi-stage approach to provenance as data and information artifacts progress along processing pipelines, is that both the quality and quantitative measures of uncertainty may be vastly improved when treated in an end-to-end manner. This also reduces the likelihood that critical information is left bahind or obscurely represented, making the later use of impossible and wasting the time and effort of all involved.

We present the current stages of implementation of our provenance infrastructure, tools and impact on what users are able to learn from the annotated information streams.