Recognizing the increased need for knowledge provenance in interdisciplinary eScience efforts, we have begun an effort to enhance a real-world data production pipeline and the resulting data services with semantic provenance. This work designing and implementing in an existing fielded virtual observatory setting has allowed us to collect key provenance requirements for a broad variety of end users. We have documented several image data pipelines for solar physics instruments at the Mauna Loa Solar
Observatory and have documented almost 20 use cases covering usage from instrument scientists, observers, data analysts and managers, and end-user scientists. These use cases have guided our work developing an initial infrastructure that can be searched, queried, or browsed by these users. We use a multi-stage approach to provenance as data and information artifacts progress along processing pipelines. Our motivation, is that both the qualitative 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 behind or obscurely represented, making the later use of the data and information difficult or impossible. Another motivation is that provenance captured consistently at ingest time supports transparency of sources and propagation of credit for data generation, thereby increasing the likelihood of contribution and reuse. 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. The Semantic Provenance Capture in Data Ingest Systems (SPCDIS) project is an NSF/OCI/SDCI funded effort involving the High Altitude Observatory at NCAR, McGuinness Associates and the University of Michigan.

DE: 9810 New fields (not classifiable under other headings)
DE: 9820 Techniques applicable in three or more fields
SC: Earth and Space Informatics [IN]
MN: 2008 Fall Meeting