Knowledge Provenance for Virtual Observatories (spcdis.hao.ucar.edu)

Peter Fox[1], Deborah McGuiness[2,3], Paulo Pinheiro da Silva[4], Stephan Zednik[1], Jose Garcia[1], Li Ding[3], Nicholas Del Rio[4], Cynthia Chang[3]

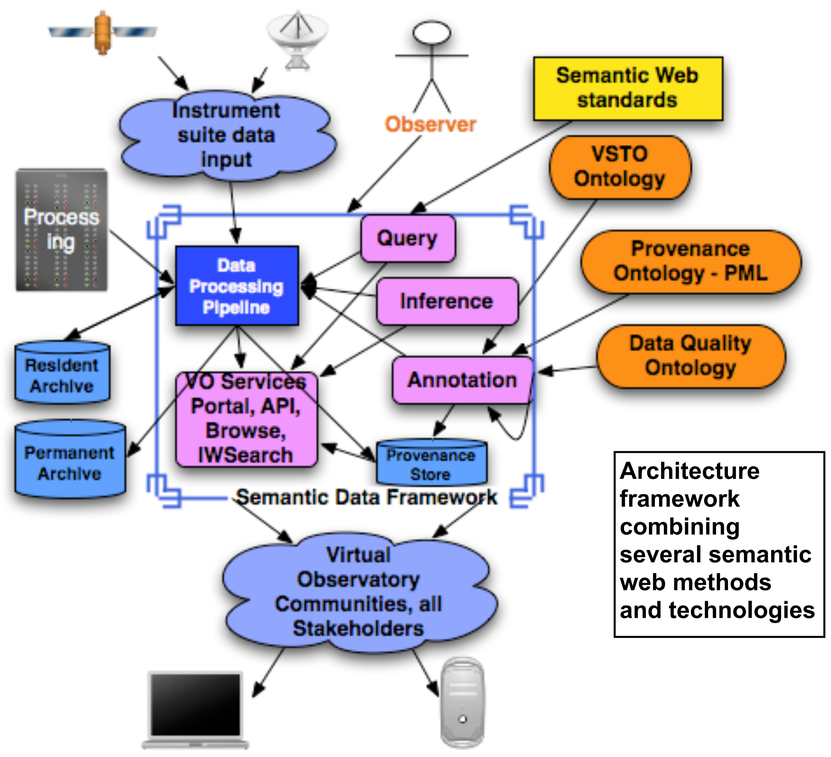
Funding from: DARPA, NSF/OCI(**OCI-0721943**) and NSF CyberShARE (**HRD-0734825**), NASA, IARPA, ARL, Lockheed Martin, Fujitsu, SRI, IBM

Tetherless World
http://tw.rpi.edu/

Background

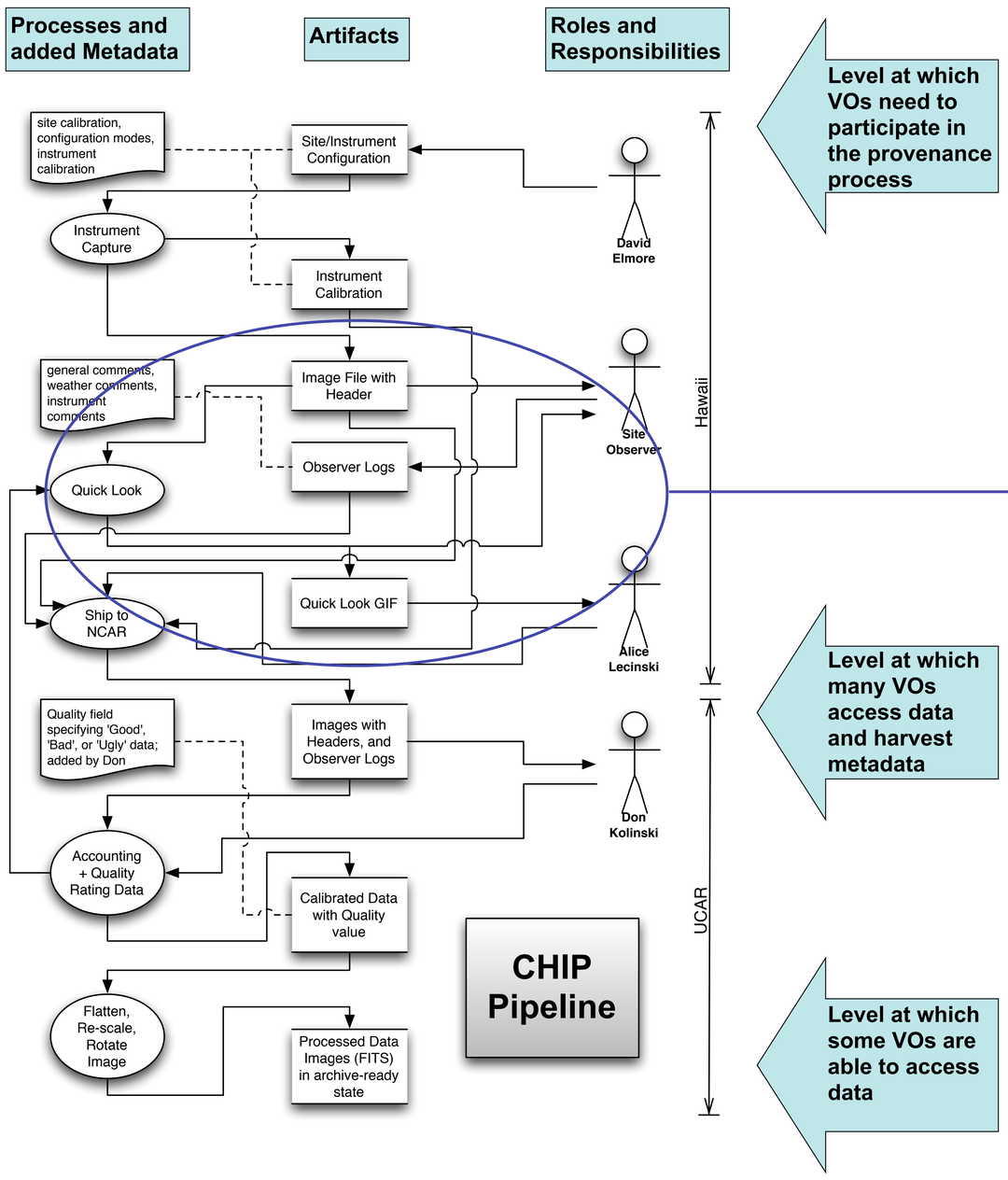
- Virtual Observatories (VOs) present a growing trend in supporting distributed interdisciplinary scientific research.
- Knowledge provenance is becoming an increasingly important component in such systems.
- Knowledge provenance includes: sources of raw data, experiments used to generate data, processing applied to the data, etc.
- We demonstrate our current progress within one existing virtual observatory -- the Virtual Solar-Terrestrial Observatory.

Architecture of a Next Generation VO

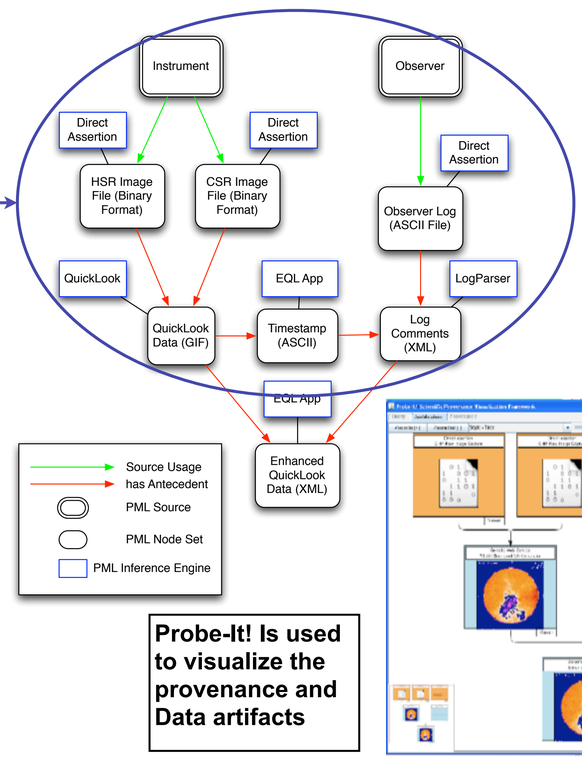


Architecture framework combining several semantic web methods and technologies

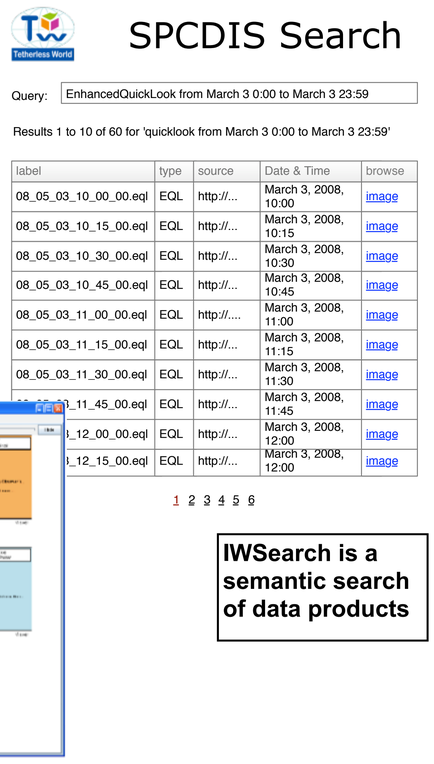
A Typical Data Ingest Pipeline



Tools to Search/Browse Provenance



Probe-It! Is used to visualize the provenance and Data artifacts



IWSearch is a semantic search of data products

- The proof markup language (PML) provides an interlingua for capturing provenance.
- The inference web toolkit provides a suite of tools for manipulating, presenting, analyzing, and searching PML in an effort to increase user understanding of information.

Motivation from Data Providers and Analysts

- Data is coming in faster and in greater volumes, outstripping conventional human quality control
- Data is being used by new communities that do not have sufficient knowledge of the data processing stages.
- Manually generated information is rarely captured and propagated in the data pipeline.
- Event determination is generally performed on the final data product, after provenance information has been lost.

Benefits/Next Steps

- Use of provenance (PML) to integrate human comments into machine-driven processes/workflows
- Multiples views of presented provenance built on same infrastructure
- Addition of structured presentation of results, answers use cases
- Next: application to science image and engineering QC use cases

SPCDIS - Semantic Provenance Capture in Data Ingest Systems

CHIP - Coronal Helium I Imaging Photometer. This instrument is operated by HAO/NCAR in Mauna Loa, Hawaii and is our testbed data ingest for provenance capture http://mlso.hao.ucar.edu/

IWSearch - Provenance Search http://onto.rpi.edu/iwsearch/

Probe-It! - A browser suited to graphically rendering Proof Markup Language (PML) based provenance associated with results derived from inference engines and workflows. http://trust.utep.edu/probeit/

Conclusion

This project supports knowledge provenance capture using semantic technologies upon ingest of data (instead of requiring hand annotation of data). Initial prototype implementation is being performed in the Virtual Solar-Terrestrial Observatory Platform.

([1]HAO/NCAR) ([2]McGuinness Associates)
([3]Rensselaer Polytechnic Inst.)
([4]University of Texas at El Paso)

NCAR · Rensselaer · Inference Web (Semantic Web infrastructure for provenance and justification) · UTEP