# Provenance-Aware Faceted Search

Deborah L. McGuinness[1,2] (dlm@cs.rpi.edu), **Peter Fox**[1] (pfox@cs.rpi.edu), **Cynthia Chang**[1] (csc@cs.rpi.edu), **Li Ding**[1] (dingl@rpi.edu), **Zhenning Shangguan**[1] (shangz@rpi.edu), **Patrick West**[1] (westp@rpi.edu), and **Stephan Zednik**[1] (zednis@rpi.edu) ([1]Rensselaer Polytechnic Institute, Troy, NY), ([2]McGuinness Associates, Latham, NY)

Abstract

Scientific information systems are expanding in many ways including providing access to more diverse information from potentially many sources. As informatics systems become more pervasive and simultaneously more diverse, the need for more sophisticated search capabilities grows. We will briefly introduce some search requirements that arise from our work on multiple interdisciplinary virtual observatory settings. We have gathered requirements for faceted search and have implemented faceted search in a science settings using a few different technical infrastructures. We will present use cases where search that is informed by provenance information is critical. We will describe some initial implementations and also highlight some more advanced provenance-aware search features including customizable facets, facet negation, and facets as views.
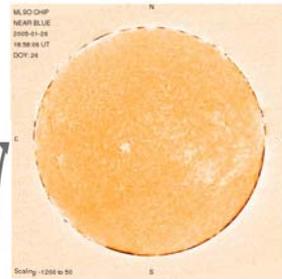
Current Guided Search
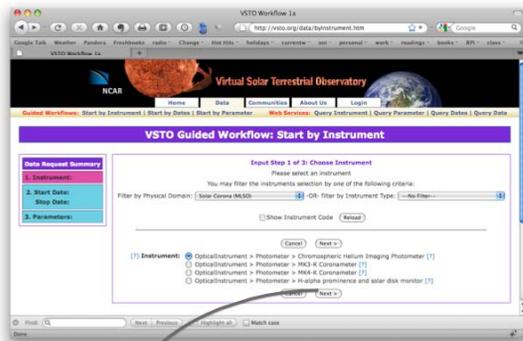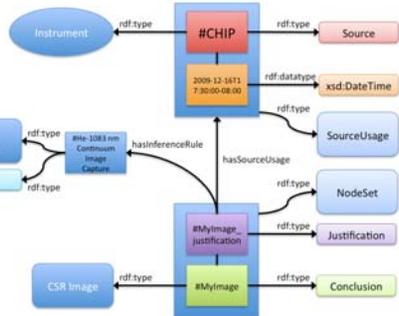

Instrument Selection



Date Selection

Parameter Selection

Data Product Selection

Resulting Image, no further information provided

The current guided search, shown on the left, is a hard-coded search within the Java code of the VSTO data portal. It guides the user through a set of web-based screens, allowing them to make selections, resulting in the display of the image (without provenance about how the image was generated and what data the image depends on).

Use ontology constructs (rdfs:subClassOf, rdfs:subPropertyOf, owl:allValuesFrom) to provide a (1) mapping between the Virtual Solar Terrestrial Observatory (VSTO) domain ontology and the provenance ontology: the Proof Markup Language (PML) and (2) domain-specific provenance augmentations to the provenance ontology. The mapping and augmentations can be in separate ontologies rather than requiring a direct update to either ontology, so they can be independent of either model.
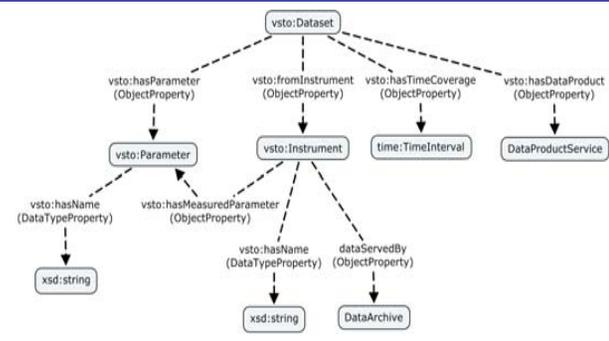
Partial VSTO domain ontology used for the current guided search, and used for the new faceted search.

Faceted Search Use-Cases

VSTO Use-Case 2 – VSTO-MLSO2: Search for and find a specific type of data in the Mauna Loa Solar Observatory database across a range of data from the Advanced Coronal Observing System. At present a scientific user needs to know a lot about the datasets, product types and processing steps to be able to identify, retrieve and use solar images. This use-case will demonstrate how ontologies, and semantically-enabled interfaces can significantly reduce the level of detail that a person has to know about the datasets and products.

Provenance Use Case – PML-MLSO2: given the result of an MLSO image from the CHIP Photometer in the information panel of a faceted search, display provenance information about source dependencies and collection information such as the atmospheric seeing conditions during the time that this data was collected at the observation point.

Same Search with Faceted Browse
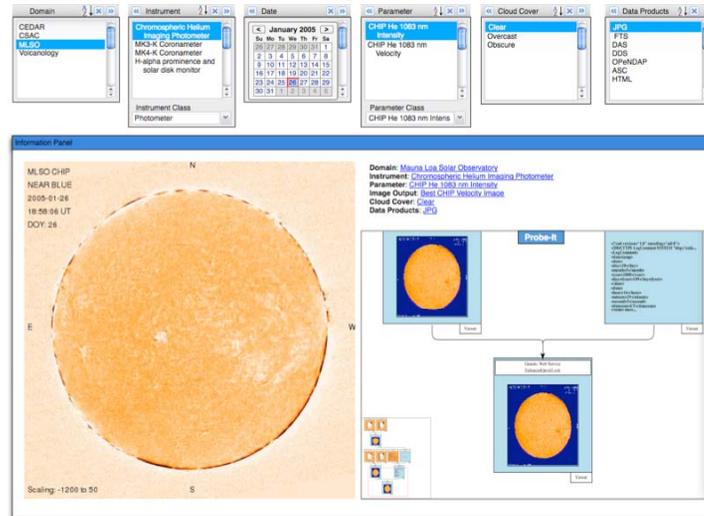
The new faceted search/browse interface allows the user to define the facets to use in the search and and provides additional facets. The Instrument, Date, parameter and Data products facets are taken from the domain ontology whereas the Cloud Cover facet is taken from the extended provenance ontology.

The Information Panel provides the user with the results of their search, retrieving the data as a JPG image. It also provides the user with information about their facet selections and an interface to browse the provenance of the resulting image, which gives the user information about data capture, observation conditions, and data calibrations.

The image tree uses the Inference Web explanation infrastructure and the Probe-It browser to display data product dependencies and provenance information.

Glossary:
VSTO - Virtual Solar-Terrestrial Observatory
MLSO – Mauna Loa Solar Observatory
CEDAR - Coupled Energetics and Dynamics of Atmospheric Research
PML – Provenance Markup Language http://inference-web.org