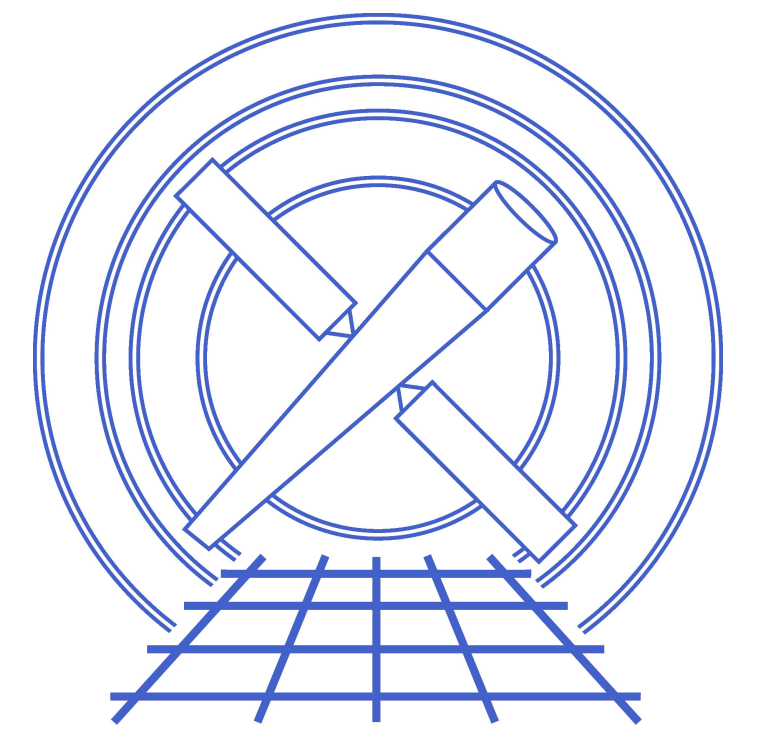


The Chandra Observational Ontology: Tying the Threads Together

S. Winkelman and A. Rots

CXC/Smithsonian Astrophysical Observatory

wink@head.cfa.harvard.edu



Introduction

Since early in the mission the Chandra Data Archive (CDA) has been providing links between ObsIds and astronomical publications to the Astrophysics Data System (ADS) as a service of our Chandra Bibliography. In the past year the CDA and ADS have embarked on an observation ontology project which has expanded our exchange of information to include details about the observations, proposals, and bibliographic classifications related to the data in publications.

ADS and the Virtual Astronomical Observatory (VAO) are developing a semantic literature search tool using the VAO base, publication, and observational ontologies to describe observations from multiple missions such as Chandra and MAST and publications associated with these missions. These descriptions are ingested into a semantic database. The descriptions are then indexed using the SOLR indexing server, and made available to astronomers using the VAO AstroExplorer. A prototype of this application is at <http://adslabs.org/semantic/alpha/explorer/publications> and a video describing some of its features and philosophy is at http://www.youtube.com/watch?v=IgFU_ehtBkU.

Individual observatories can contribute as much as they are able to for this project and information exchange formats are so far being handled on an observatory-by-observatory basis.

The Chandra Observational Ontology

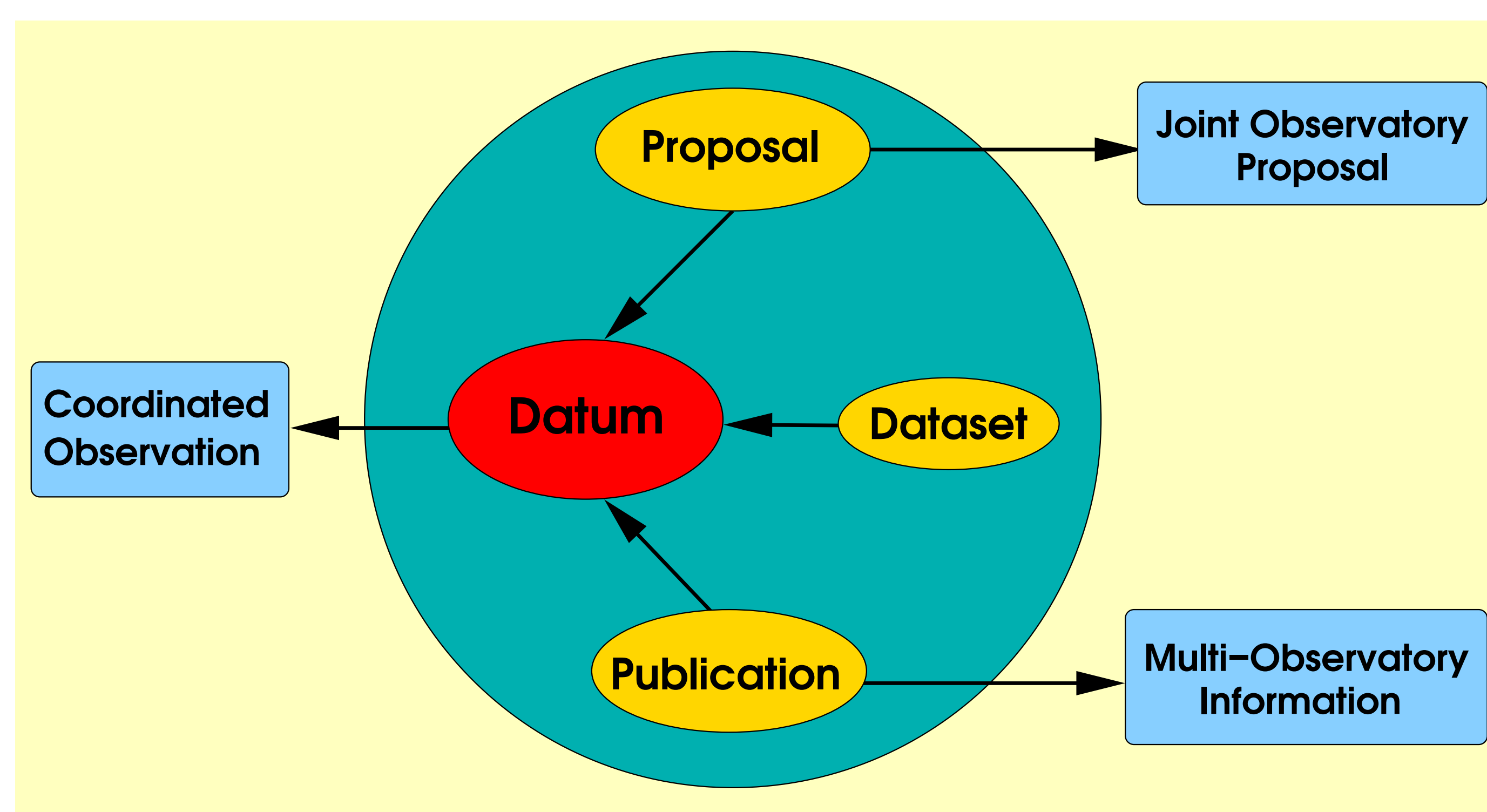


Figure 1: The Chandra ontology is focused on the Datum or ObsId. Information within the cyan circle is controlled/collected by the CDA. The gold ellipses represent categories of information linked to the Datum by the CDA. The blue boxes are information links to other observatories.

The Chandra Observational Ontology provides a small part of the entire ontology. The Datum or ObsId is at the center of the Chandra ontology. Proposals, Datasets, and Publications are then linked to a Datum, providing a richer picture of what a Chandra observation is. The next step is to link to information from other observatories to create an even more complete picture of Chandra observations.

Datum: A Datum arises from a single simple observation and could be a spectrum, image, cube, etc. The key is that a Datum is one set of photons. For the CDA this is known as an ObsId. A Datum is described by a number of attributes including target name, coordinates, instrument configuration, observation date, etc.

Dataset: A Dataset is more complex. Examples are monitoring series or grid-type observations where multiple observations of the same target are observed to accomplish a scientific goal. For Chandra, these tend to be aggregates of ObsIds.

Proposal: ObsIds (the Datum) are spawned from Proposals. A Proposal includes such things as title, abstract, principal investigator, total time awarded, and when the award was made.

Publication Data are published in papers. Many observatories, including the CDA, link observations to publications while maintaining their bibliographies. In addition to links to data, a publication's metadata includes when the paper was classified and any flags which have been added to more fully describe the data analysis contained within the paper as it relates to the observatory.

Links Beyond the Observatory: Three obvious extra-observatory links have been identified in the Chandra ontology: joint proposals where Chandra can grant time at another observatory and vice versa; coordinated observations between Chandra and other observatories; and published multi-observatory data analyses.

Bibliographic Implications

There are some publications in the Chandra bibliography which are clearly presenting results from analysis of Chandra data but which do not provide enough information within the paper for the curators to link to specific Chandra data. This linking is very important for our bibliography statistics, so in cases where we cannot confidently link to data, we add flags in our database so the paper is counted in our statistics in some way.

We have recently defined a new metric for our bibliography which relies on having links to the data, not just a flag. This new metric is presented in Poster P128 by Arnold Rots. As a result, our ontology will become more complete as we implement new data tracking procedures in our bibliography curation.

Acknowledgment

We thank ADS, especially Rahul Dave, for guidance on this project. This work has been supported by NASA under contract NAS8-03060 to the Smithsonian Astrophysical Observatory for operation of the Chandra X-ray Center.

Tying the Chandra Threads Together

Physically, an instance of the Chandra ontology is a set of XML files generated from Perl scripts which are stored in a public ftp area. In addition, a set of XML files are generated which provide ADS with information on changes made to the ontology so they can make updates to its semantic database.

A large portion of the information which is contained in the Chandra Observational Ontology comes from two databases which have been maintained throughout the mission: the Observation Catalog and the bibliography database. These two databases comprise all of the information contained within the *Datum*, *Proposal*, and *Publication* portions of the Chandra ontology.

In order to populate the *Dataset* portion of the ontology, we had to construct a new database which defines and links observations together based on how the observations are related. To date we have identified six types of datasets for Chandra data:

Monitor Series: a set of ObsIds which are to be observed sequentially with a defined temporal spacing;

Group: a set of ObsIds which are to be observed with a set time period;

Grid: a set of ObsIds which are spatially close together;

SeqNum: a set of ObsIds which comprise a long observation that was split for scheduling reasons;

Proposal: a set of ObsIds which are spawned from a proposal; and

Defined Dataset: a set of ObsIds which have been assigned a Custom Dataset Identifier that form a logical collection in at least one Publication.

Currently the Dataset linking is a prototype database which is maintained on the side by the Chandra archive operations team until all of the kinks and pitfalls can be worked out or defined. The next step will be to incorporate these aggregates into the Chandra Observation Catalog and make the Chandra X-ray Center (CXC) data access tools like WebChaser aggregate aware, thus allowing users to identify complete sets of data for analysis.

The links to other observatories exist in an incomplete form within the CXC. Some progress has been made to identify the current links and to tie them into the Chandra ontology. In addition, tools will be developed to help collect this additional information during the proposal and scheduling portions of an observation and link that information within the Chandra Observation Catalog.

ADS Ties It ALL Together

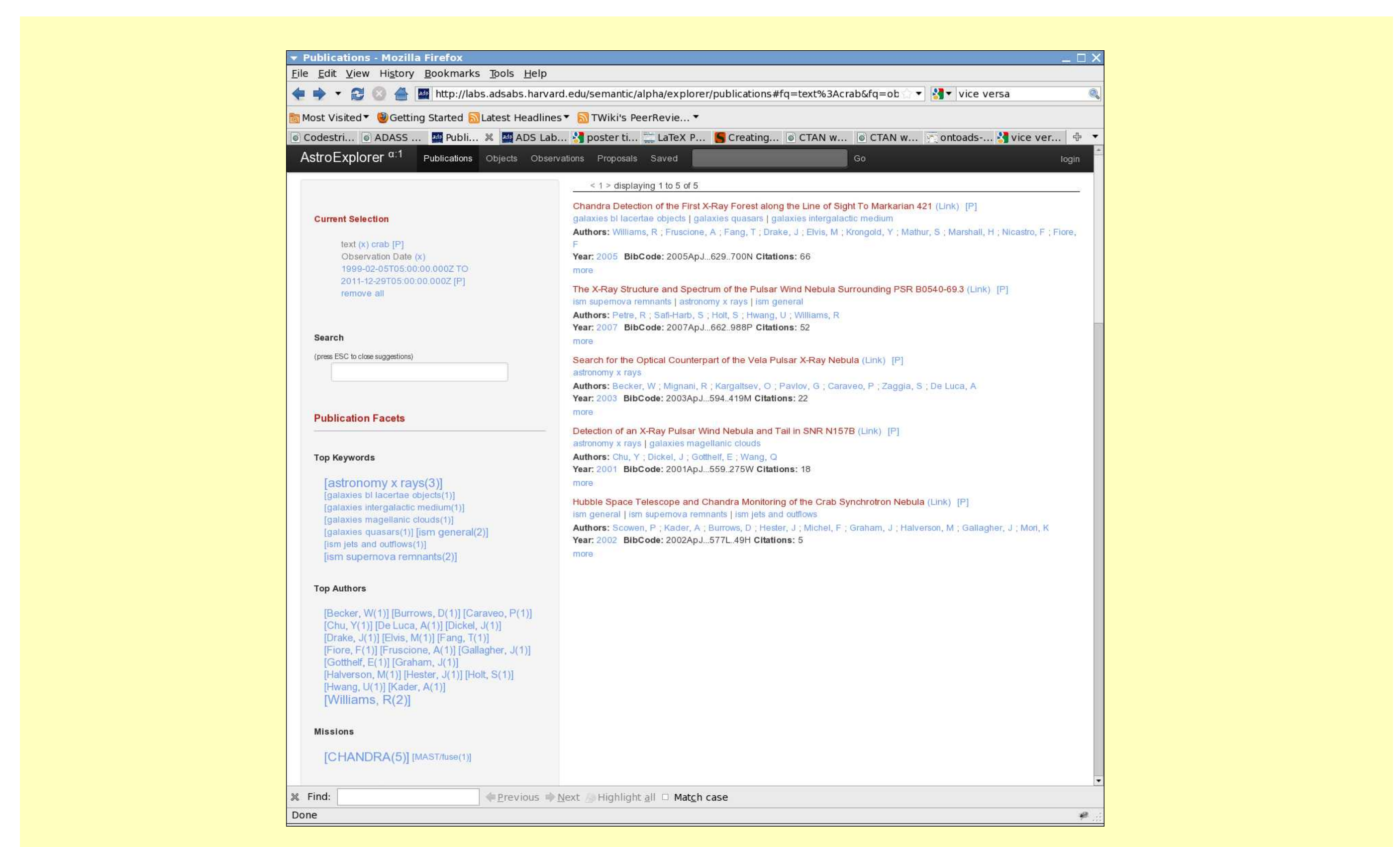


Figure 2: Snapshot of the prototype ADS semantic literature search tool.

The complete observational ontology as envisioned by ADS can encompass all of the details needed to describe the scientific process. It includes aspects which describe astronomical sources such as NED and SIMBAD sources; observatories and what telescopes/instruments they operate; proposals; science processes such as observations, experiments, simulations, and theories; science products including data products, hardware products, and publications; and much more.