OntoADS

Machine readable and interoperable descriptions of existing bibliographic efforts

http://rahuldave.github.com/ontoads

Rahul Dave
rahuldave@gmail.com
What resources?

- Objects
- Papers
- Datasets
- Instruments
- Collaborations
- Observatories
- Funders
- Curators and Agents
- Proposals and Grants

RESEARCH and the PROCESS of research
How do we work?

- We put in proposals and hope we get grants.
- We use the grants to support ourselves, students, post-docs, etc.
- We do research, using data taken by others or us
- at Observatories using Instruments like satellites and telescopes, having made ObservationProposals and having been granted time (sim for labs).
- We then write papers, where we cite prior art in the field,
- and Repeat and Rinse....,
- producing new sets of Observations and Data, from which we deduce the existence of certain Objects.
High-level Data Products

Mission Design

Observing Proposals

Simulations

Scientific Lifecycle

Analysis

Observations

Data Reduction

Pepe et al, JASIST (2009)
PROCESS: What else?

- aggregation and exploration of resources
- recommendation, search, and discovery of resources
- the creation of metrics to measure our work
- attribution, provenance and repeatability, preservation and versioning.
- PROFIT! ....errr....

“O.K. let’s slowly lower in the grant money.”
Todd Bearson
Arlington, Mass.
What tech?

- Semantic Web and Linked Data
  - resources are named via HTTP URIs
  - metadata is *open* and in a standard format
  - links between resources are *typed*
  - built on the architecture of the web, *no APIs*
- RDF for Triples and OWL-DL for Ontology
- Triplestore and Database
- Data Entry via machine learning and curation
Why Semantics

<http://www.cambridgesemantics.com/people/about/lee> foaf:name "Lee Feigenbaum".

OWL

:Parent owl:equivalentClass [
  rdf:type owl:Restriction ;
  owl:onProperty :hasChild ;
  owl:someValuesFrom :Person ] .

RDF

● a common language for resources and links, which can:
  ○ be queried in standard fashions using off-the-shelf technology based on standards (SPARQL)
  ○ be used to make blame-annotated statements about the links themselves (REIFICATION), via curation or by the results of, for e.g., machine learning algorithms.
  ○ such that descriptions of resources and statements can be federated across organizations like ESO, STSci, ADS
● Enable us to become part of the global web of linked data, buying us external indexing for free.
● Inferencing, Inferencing, Inferencing.
What Applications?

LEVERAGE OTHERS' WORK

- faceted exploration
- in a research folder or portfolio
- with links for education and
- with searches and exploration
- with results of data mining as recommendations
- and inferences from the corpus of data to answer
- arbitrary questions we haven't thought of
- and to establish provenance for research and curation
- also metrics for publishers, funders, agencies
- and support for the authoring and curating process
- all done in a way that's part of the linked-data web
- so that we may be indexed
Soon

Facets

SAO/NASA Astrophysics Data System (ADS)

Query Results from the ADS Database

Selected and retrieved 200 abstracts.

<table>
<thead>
<tr>
<th>#</th>
<th>Bibcode</th>
<th>Authors</th>
<th>Score</th>
<th>Date</th>
<th>Title</th>
<th>List of Links</th>
<th>Access Control</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2006glsw.book..269S</td>
<td>Schneider, P.</td>
<td>81.000</td>
<td>n/a 2006</td>
<td>Weak Gravitational Lensing</td>
<td>A F X</td>
<td>R C e</td>
<td>U</td>
</tr>
<tr>
<td>2</td>
<td>2008PhR...462...67M</td>
<td>Munshi, Dipak; Valageas, Patrick; van Waerbeke, Ludovic; Heavens, Alan</td>
<td>68.000</td>
<td>Jun 2008</td>
<td>Cosmology with weak lensing surveys</td>
<td>A F X</td>
<td>R C e</td>
<td>U</td>
</tr>
<tr>
<td>3</td>
<td>2003ARA&amp;A...41...645R</td>
<td>Refregier, Alexandre</td>
<td>61.000</td>
<td>n/a 2003</td>
<td>Weak Gravitational Lensing by Large-Scale Structure</td>
<td>A F X</td>
<td>R C e</td>
<td>U H</td>
</tr>
<tr>
<td>4</td>
<td>2008ARNPS..48...99H</td>
<td>Hoekstra, Henk; Jain, Bhuvenesh</td>
<td>51.000</td>
<td>Nov 2008</td>
<td>Weak Gravitational Lensing and Its Cosmological Applications</td>
<td>A X</td>
<td>R C e</td>
<td>U</td>
</tr>
<tr>
<td>5</td>
<td>2003astro.ph..6465S</td>
<td>Schneider, Peter</td>
<td>44.000</td>
<td>Jun 2003</td>
<td>Gravitational lensing as a probe of structure</td>
<td>A X</td>
<td>R C e</td>
<td>U H</td>
</tr>
<tr>
<td>6</td>
<td>2006MNRAS.368.1323H</td>
<td>Heymans, Catherine; Van Waerbeke, Ludovic; Bacon, David; Bacchini, J. Bonavera</td>
<td>41.000</td>
<td>May 2006</td>
<td>The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations</td>
<td>A G E</td>
<td>R C e</td>
<td>U</td>
</tr>
</tbody>
</table>

faceted browsing
Curation after mining

The neural stem cell microenvironment

In mammals, neural stem cells appear early in development and remain active within the central nervous system for the whole life and reside within changing microenvironments whilst retaining the basic properties of a stem cell: multipotentiality and the ability to divide along with the fundamental structural components and signaling molecules of their microenvironments. In early neural development, they are situated among other neuroepithelial cells and they are exposed to various signals such as retinoic acid, sonic hedgehog, and the complexity of their microenvironment increases due to the emergence of various types of neuronal progenitor cells. The morphology and reside in specific microenvironments that are called neurogenic niches; small neurogenic islands are present in the adult brain and under the control of microenvironmental cues.

1. The embryonic neural stem cell (NSC) microenvironment

Central nervous system (CNS) development is an intricate process relying on a series of mechanisms precisely regulated in time and space to achieve the correct location and maturation of neural progenitors towards their respective destination within an approximately one-week period during embryogenesis. The embryonic CNS is a dynamic structure, constantly increasing in size due to neurogenesis, while the stem
• SPARQL

PREFIX space: <http://purl.org/net/schemas/space/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT *
{ ?launch space:launched ?date
  FILTER ( ?date > "1968-10-1"^^xsd:date && ?date < "1968-10-30"^^xsd:date )
}

• Inference
• No API, just queries
• Anyone can build

Lots of apps will be scripts. Most apps are as yet unthought of!

Long running scripts will be used by curators, funders, program directors to realize and federate metrics.

Post inference indexes will be built for faster retrieval.

OPEN METADATA. OPEN FORMATS.
Why here and what?

**ADS**

- as part of VAO efforts
- have access to large bib database
- relationships and prior art from Survey Librarians for ESO, Chandra, NRAO, STSci
- links to NED, SIMBAD
- Build on ontologies from [SWAN](#) at MGH, DC, etc!

**Ontologies:**

Base, Keywords(SKOS): AAKEYS(IVOA), Biblio, Observations/Experiments, Objects.

*Others ought to be able to build on this!*
- Person
- PersonName
- Agency, FundingAgency
- ObsvAgency, LabAgency
- Collaboration, Membership
- Organization, Affiliation
- Agent
- Software
- Observatory, Lab
- Proposal, ObsvProposal, LabProposal
- Program, ObsvProgram, LabProgram
- Instrument
Publications

- Authors
- Keywords
- Affiliations
- Citations
- Links to data, objects, proposals.
- Readership
_9:nbachall a agent:PersonName;
foaf:name "Bahcall, N".

_9:rsoneira a foaf:Person;
   foaf:name "Soneira, R".

   ads:bibcode "1983ApJ...270...20B";
   adsbib:bibcode "1983ApJ...270...20B";
   adsbib:author _9:nbachall, _9:rsoneira;
   dc:title "The spatial correlation function of rich clusters of galaxies";
   dc:subject aakeys:AngularCorrelation, aakeys:MethodsDataAnalysis;
   bibo:abstract "A series of objective statistical estimators is applied to directly study the three-dimensional distribution.

   <http://adset.cfa.harvard.edu:8080/1957AJ.....62..248S>,
   <http://adset.cfa.harvard.edu:8080/1958ApJS....3..211A>,

   ...........
   bibo:pageEnd "38";
   bibo:pageStart "20";
   bibo:volume "270".
Curation by Humans or Agents!

- Paolo Ciccarese
- John Doe

- foaf:Person
- foaf:name
- reification:subject
- reification:object
- is-a
- agents:ReifiedAka
- agents:PersonName
- agents:fullName
- pav:createdOn
- pav:createdBy

- October 30th 2008

Colors:
- FOAF
- Provenance, Authoring and Versioning
- Reification
Observations

- Unique Dataset Identifiers: archives, datasets, data-products, images.
- Position, footprint and date
- Objects observed
- Wavelengths
- Instruments
- People, Proposals, Collaborations
Objects

- object-type
- catalogs and catalog identifiers
- ra, dec, survey-field, footprint, survey
- spectrum, light-curve, observation times
- multi-wavelength coverage
- GCVS variability classification, algorithms
- NED/SIMBAD identifiers, data model
- papers, datasets, images
taosdb:taos-paper-on-variables
adsbib:authoredByCollaboration surveysdb:TAOSCOLLABORATION
adsbib:asAResultOfProposal :NSF-1002;
adsobsv:hasObservation :TAOS-F152-R7.

:TAOS-F152-R7 atObservatory taosdb:LulinObservatory.

:thisclassification a gcvs:Classification;
  adsobject:classifiedIn taosdb:taos-paper-on-variables;
  adsobject:classList gcvs:Cepheid.

:TAOS-152/001 a adsobject:Star;
  adsobject:classifiedAs :thisclassification.

:TAOS-152/001 owl:sameAs usnodb:USNO-1657-0005791.

-----

Somewhere else:

-----

_10:USNO-1657-0005791 adsobject:classifiedAs [ a gcvs:Classification;
  adsobject:classifiedIn machodb:someotherpaper;

-----

oops. The system can detect such oops by *Inferencing*. 
Pie in the Sky (but more on ground)

Currently Working On: Base, SKOS, Biblio Ontologies. With Test Cases. (May End)

Soon Working On: Object, Obsv/Exp Onts, Scripts (Jul End)

Software Working On: Faceted Browser for Pubs/RDF (Aug End)

- Bio like claim and protocol provenance
- inference inference inference
- metadata vs data (usenet)
- annotation?
- direct data access. direct visualization.
- enhanced journals
- pave the way to orcid
- possibly required linked data compatibility (data.gov)
- metadata to describe aggregates(ORE)
- interlinking as part of peer-review process?
HELP!


No standards body yet. But we need input. Especially Use cases.

Especially need help from HEP Community.

Go to ontoads on github and help us out. Make comments. Fork the ontologies and play. Tell us if they serve your purpose. And where not.

Get your voice in.
http://rahuldave.github.com/ontoads

Thanks to: ADS. SWAN. MSR. MGH.