

OntoADS

Machine readable and
interoperable descriptions of
existing bibliographic efforts

<http://rahuldave.github.com/ontoads>

Rahul Dave
rahuldave@gmail.com

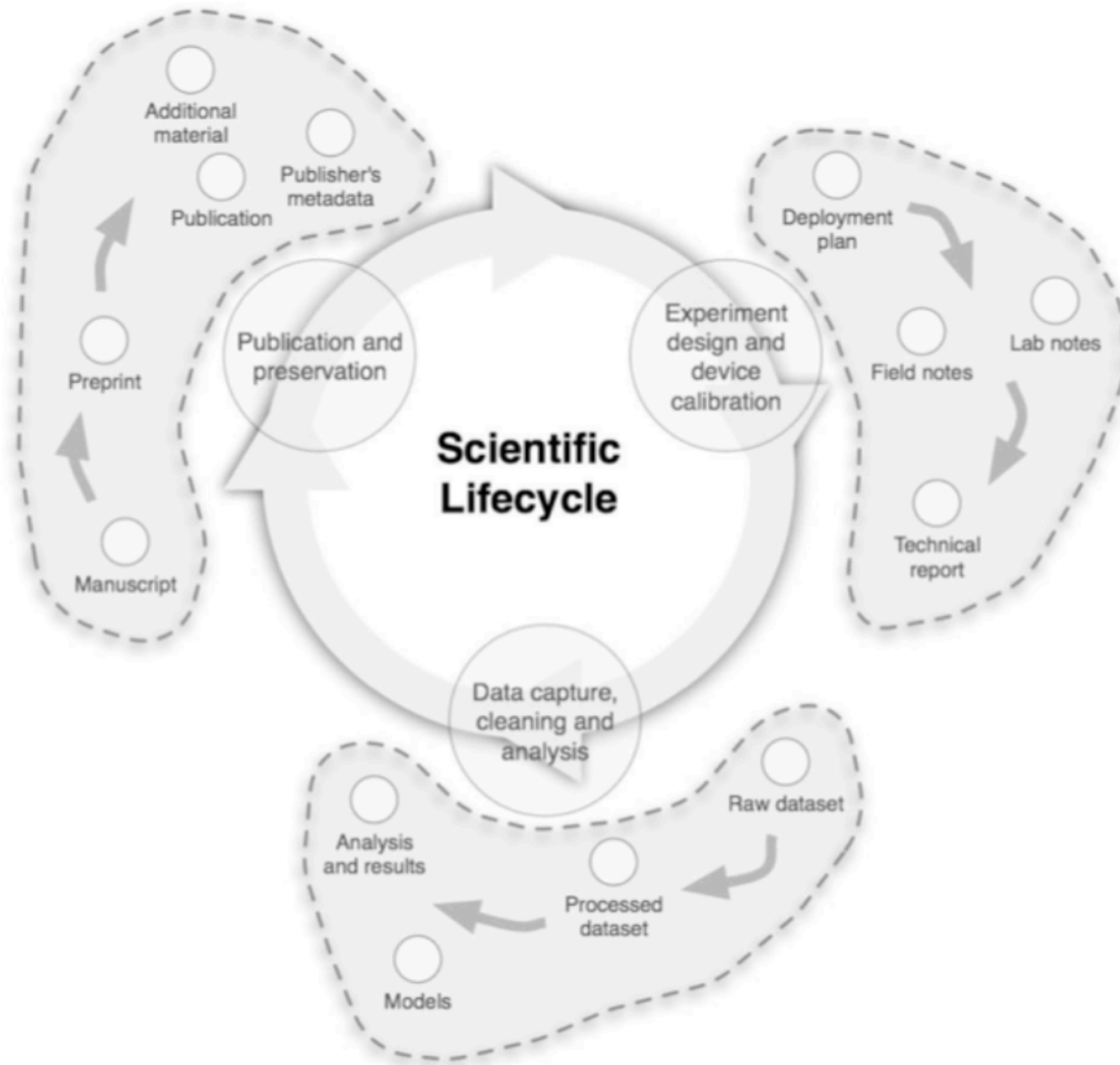
What resources?

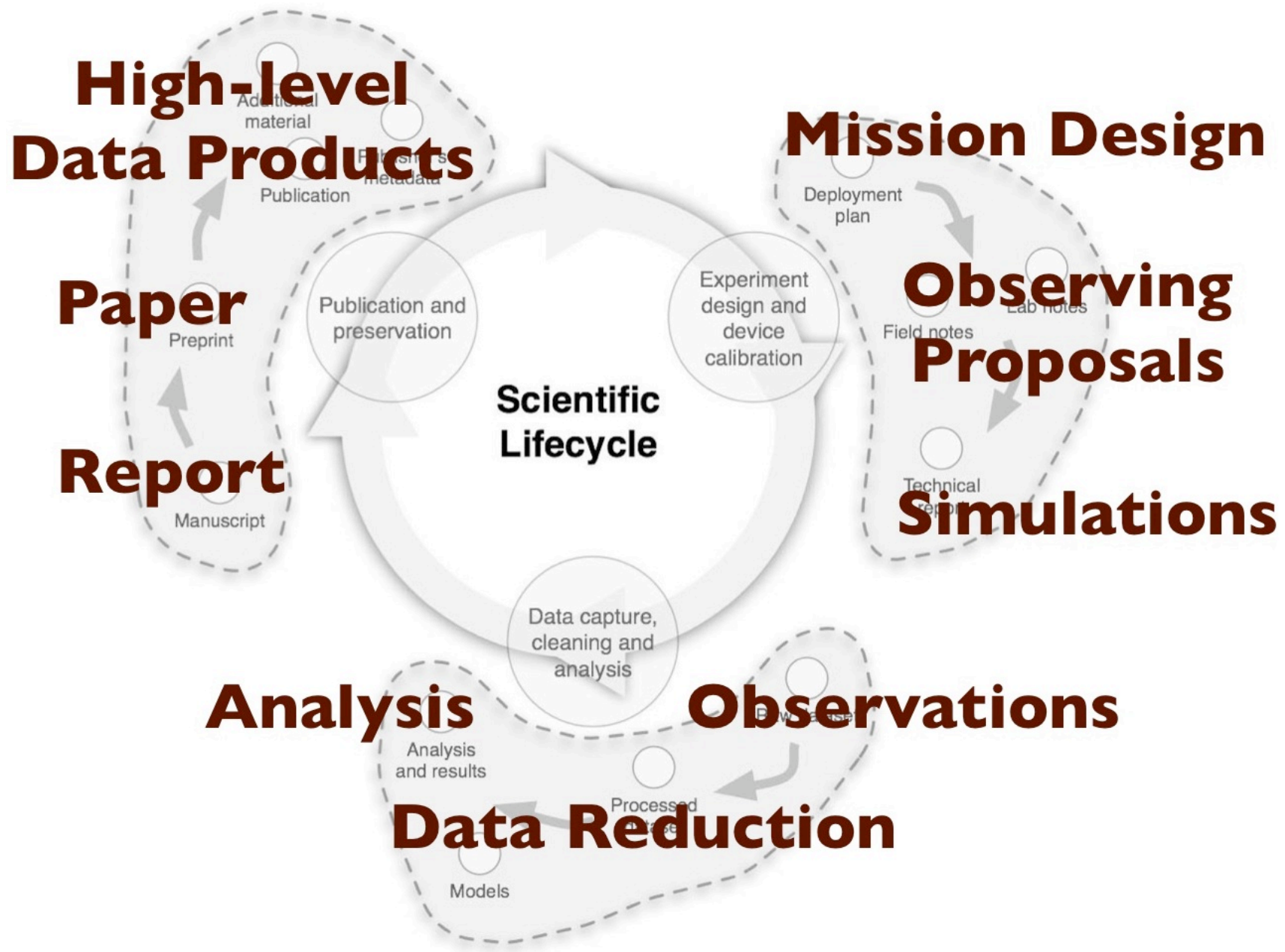
- Objects
- Papers
- Datasets
- Instruments
- Collaborations
- Observatories
- Funders
- Curators and Agents
- Proposals and Grants

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..

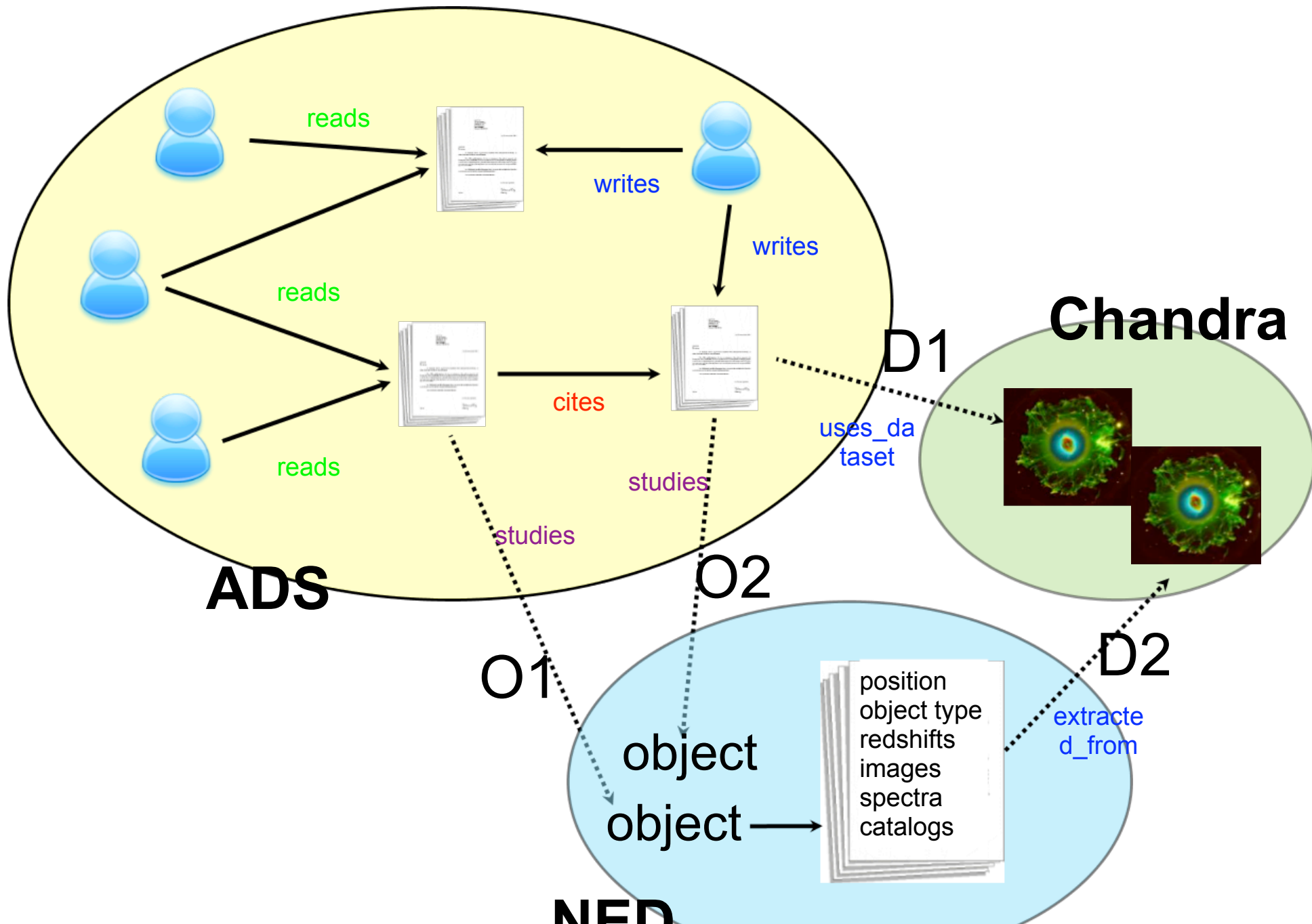




Pepe et al, JASIST (2009)

Links?

ADS/NED/Chandra



PROCESS: What else?



"O.K., let's slowly lower in the grant money."

Todd Bearson
Arlington, Mass.

- 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....

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

OWL

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

RDF

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

- 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

Results

ADS Query Results

http://adsres.cfa.harvard.edu/cgi-bin/topicFacetSearch?q=%22weak%20lensing%22;qtype=INSTRUMENT

Google

Apple Yahoo! Google Maps YouTube Wikipedia News (284) Popular

SAO/NASA Astrophysics Data System (ADS)

Query Results from the ADS Database

[Go to bottom of page](#)

Related Objects

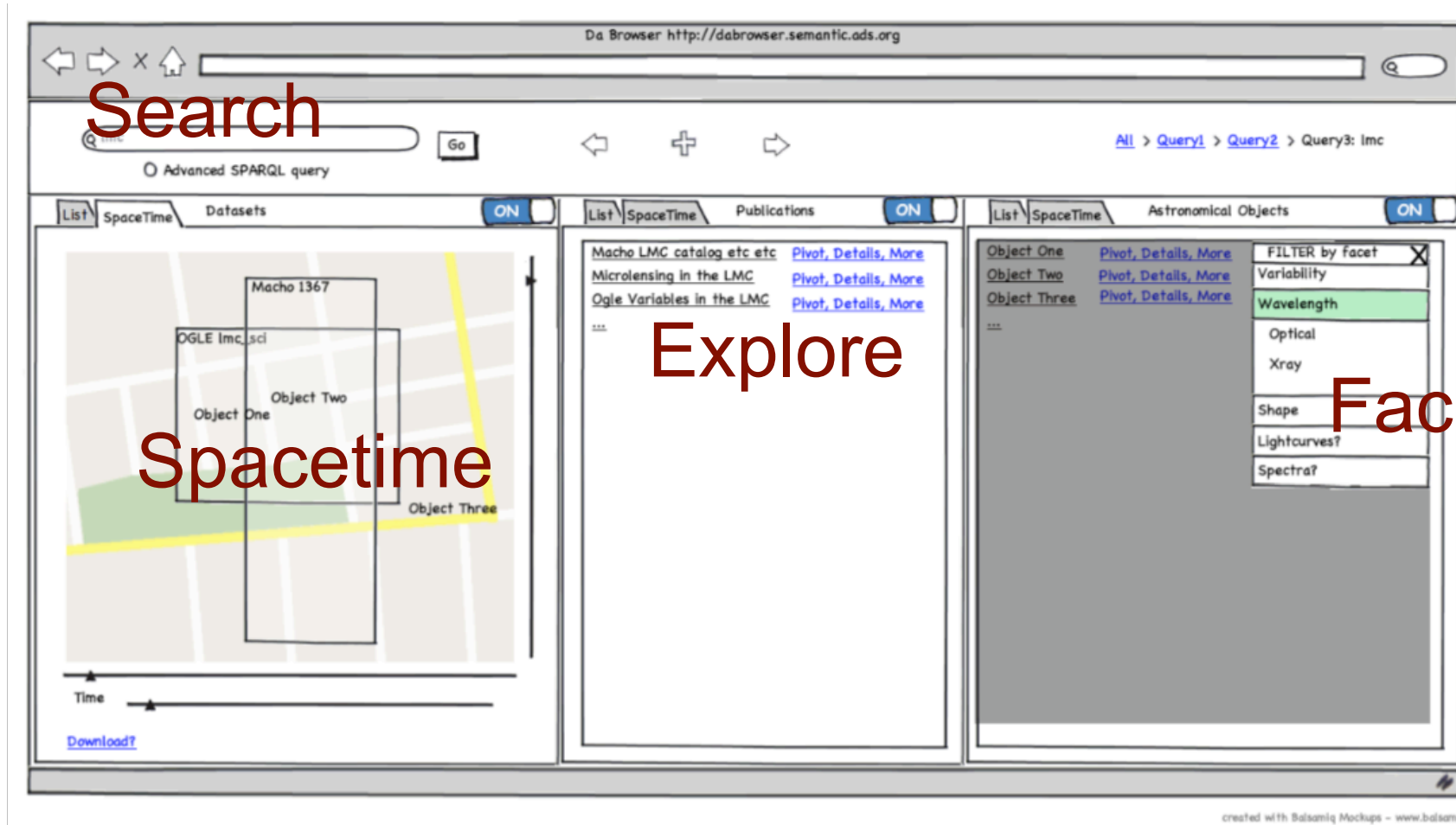
[NAME HDF \(12\)](#)
[NAME Chandra Deep Field-South \(8\)](#)
[ACO 1689 \(8\)](#)
[ACO 901 \(8\)](#)
[ZwCl 0024+1652 \(7\)](#)
[ACO 902 \(7\)](#)
[ACO 2390 \(6\)](#)
[ZwCl 1455+2232 \(5\)](#)
[NAME SGP \(4\)](#)
[CIG 2137.3-2353 \(4\)](#)
[ACO 2261 \(4\)](#)
[ACO 2219 \(4\)](#)
[ACO 2218 \(4\)](#)
[ACO 1835 \(4\)](#)
[ACO 1763 \(4\)](#)
[ACO 370 \(4\)](#)
[ACO 267 \(4\)](#)
[ACO 209 \(4\)](#)
[ACO 68 \(4\)](#)
[1E 0302.5+1717 \(4\)](#)
[ZwCl 1358+6245 \(3\)](#)
[NAME HDF-S \(3\)](#)
[NAME COSMOS FIELD \(3\)](#)
[NAME CBL \(3\)](#)
[NAME ACO 901A \(3\)](#)
[NAME ACO 901-902](#)
[SUPERCLUSTER \(3\)](#)
[CIG 1224.7+2007 \(3\)](#)
[CIG 1054.4-0321 \(3\)](#)
[ACO 1942 \(3\)](#)
[ACO 1364 \(3\)](#)

Selected and retrieved 200 abstracts.

Sort options

#	Bibcode Authors	Score	Date	List of Links Access Control Help
1	<input type="checkbox"/> 2006glsw.book..269S Schneider, P.	81.000	n/a 2006	A E X R C c U
2	<input type="checkbox"/> 2008PhR...462...67M Munshi, Dipak; Valageas, Patrick; van Waerbeke, Ludovic; Heavens, Alan	68.000	Jun 2008	A E X R C c U
3	<input type="checkbox"/> 2003ARA&A..41..645R Refregier, Alexandre	61.000	n/a 2003	A E F X R C c U H
4	<input type="checkbox"/> 2008ARNPS...58...99H Hockstra, Henk; Jain, Bhuvnesh	51.000	Nov 2008	A X R C c U
5	<input type="checkbox"/> 2003astro.ph..6465S Schneider, Peter	44.000	Jun 2003	A X R C c U H
6	<input type="checkbox"/> 2006MNRAS.368.1323H Heymans, Catherine; Van Waerbeke, Ludovic; Bacon, David; Borgeau-Chavez, Jean-Pierre	41.000	May 2006	A E G X R C c U

faceted browsing



Data

Papers

Objects

apod

Curation after mining



SCF Annotation Editor

The neural stem cell microenvironment

Save Annotations

Original

Listings

Enter term to search for:

Terms: neurogenesis :59 ([x d a s o](#)) growth :30 ([x d a h o](#)) cell-cell signaling :28 ([x d a s o](#)). Showing matches for ONLY growth

Annotation id 6. Updated at 2009-07-20 19:47:51.020766 by Mining Robot.

- Ilias Kazanis^{1,4},
- Justin Lathia^{1,2},
- Lara Moss^{1,3},
- Charles French-Constant^{1,3}

¹Department of Pathology, University of Cambridge, Tennis Court Road, CB2 1QP, Cambridge, UK
²Current address: Dept. of Surgery, Division of Neurosurgery, Duke University Medical Center, Durham, NC 27710, USA
³Current address: MRC Centre for Regenerative Medicine, The Queen's Medical Research Institute, 47 Little France Crescent

In mammals, neural stem cells appear early in development and remain active within the central nervous system for the whole life span and reside within changing microenvironments whilst retaining the basic properties of a stem cell: multipotentiality and the ability to self-renew along with the fundamental structural components and signalling molecules of their microenvironments. In early neural development, stem cells; they are situated among other neuroepithelial cells and they are exposed to various signals such as retinoic acid, sonic hedgehog, and glial cells and the complexity of their microenvironment increases due to the emergence of various types of neuronal progenitor cells, astroglial morphology and reside in specific microenvironments that are called neurogenic niches; small neurogenic islands which are operating during embryonic development.

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 reach their respective destination within an approximately one-week period during embryogenesis. The embryonic CNS is a dynamic structure, constantly increasing in size due to histogenesis, while the stem cell population is present in the adult

Annotate

[Delete All](#) | [Add Mapping](#)
go : nervous system development [Delete](#) | [Alter](#)
go : cell growth [Delete](#) | [Alter](#)
go : growth pattern [Delete](#) | [Alter](#)
go : cell-cell signaling [Delete](#) | [Alter](#)
go : system development [Delete](#) | [Alter](#)
go : growth [Delete](#) | [Alter](#)
go : stem cell development [Delete](#) | [Alter](#)
go : cell development [Delete](#) | [Alter](#)

- 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?

*Curators and Authors
should never see any
tech! Thus our effort
should supply:
MACROS,
VALIDATION,
FORMS.*

ADS

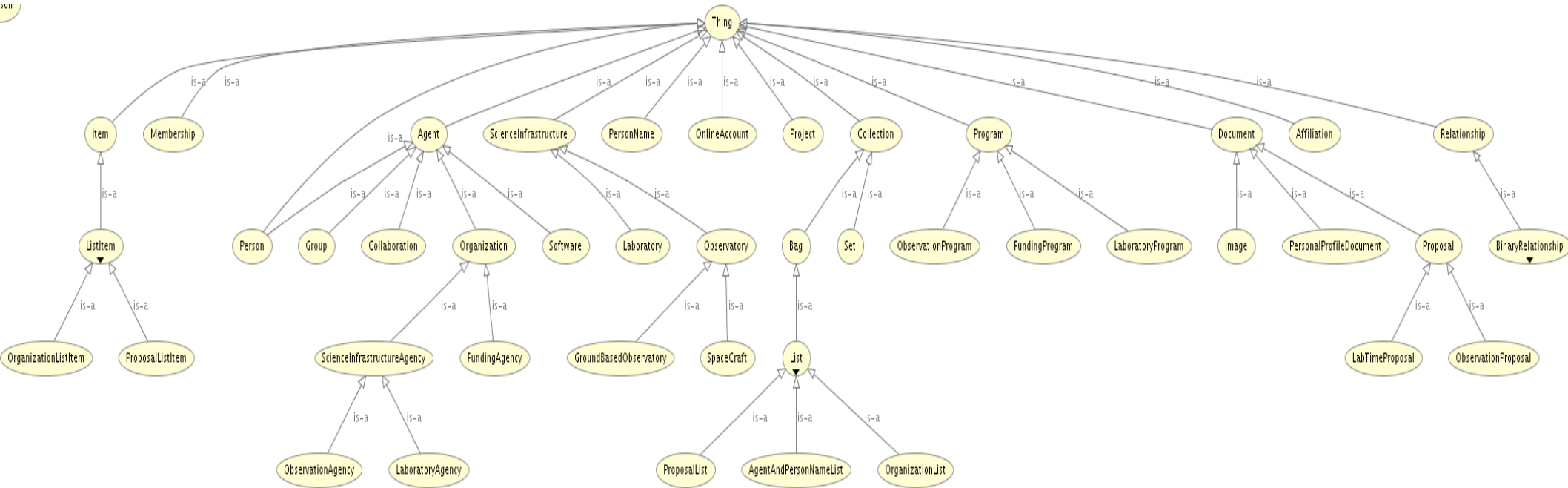
- as part of VAO efforts
- have access to large bib database
- relationships and prior art from Survey Librarians for ESO, Chandra, NRAO, STSci
- links to NED, SIMBAD
- Build on ontologies from [SWAN](#) at MGH, DC, etc!

Ontologies :

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

Others ought to be able to build on this!

SUI



- 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

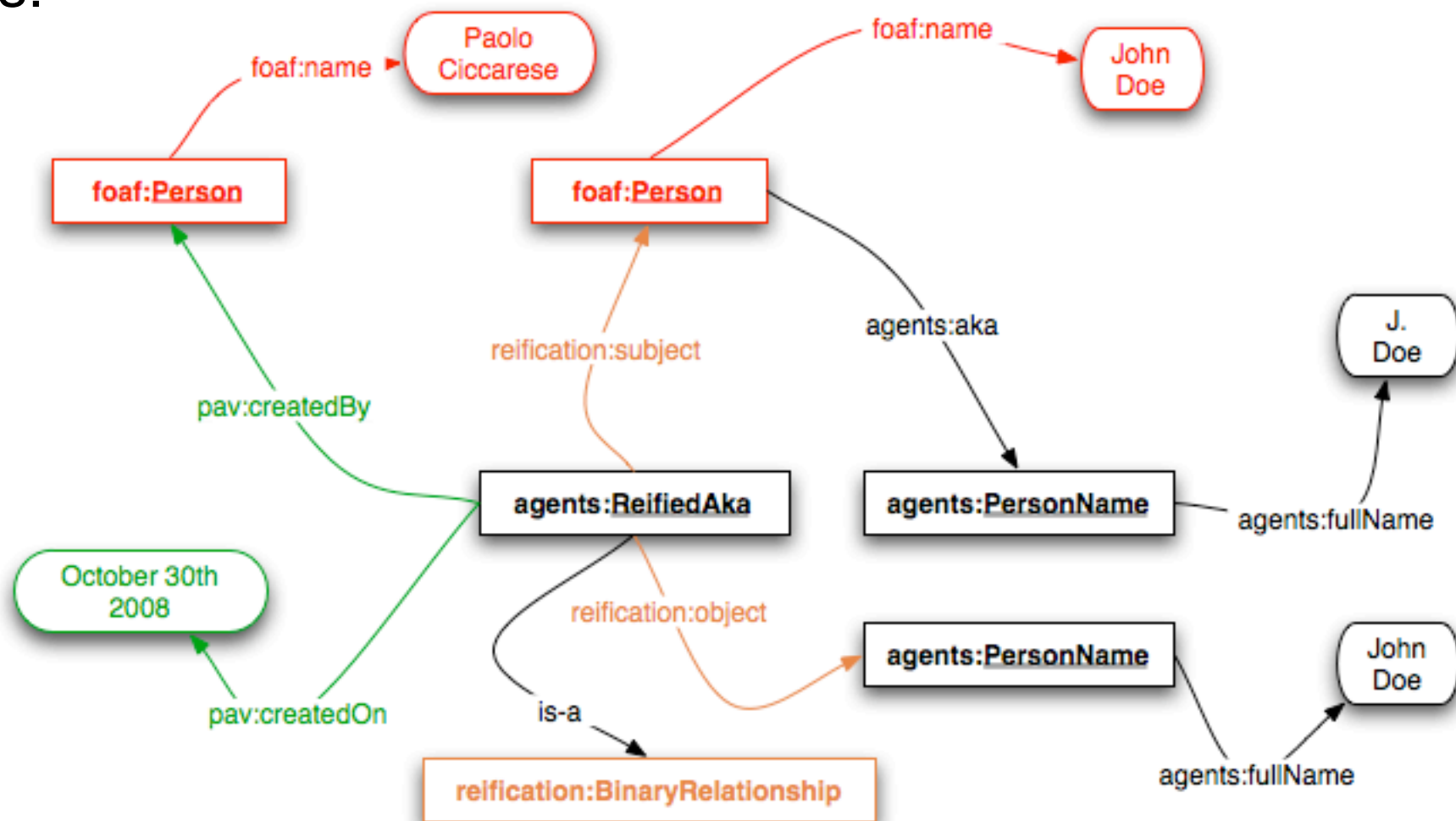
@prefix adsbase: <http://github.com/rahuldave/ontoads/owl/adsbase.owl#>.
@prefix agent: <http://swan.mindinformatics.org/ontologies/1.2/agents/>
@prefix ads: <http://github.com/rahuldave/ontoads/owl/adsbib.owl#>.
@prefix bibo: <http://purl.org/ontology/bibo/>.
@prefix dc: <http://purl.org/dc/elements/>.
@prefix dct: <http://purl.org/dc/terms/>.
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix skos: <http://www.w3.org/2004/02/skos/core#>.
@prefix aakeys: <http://github.com/rahuldave/ontoads/owl/AAKeys.rdf#>.
@prefix xml: <http://www.w3.org/XML/1998/namespace>.
@prefix _9: <http://adsset.cfa.harvard.edu:8080/>.

_9:nbachall a agent:PersonName;
foaf:name "Bahcall, N".

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

<http://adsset.cfa.harvard.edu:8080/1983ApJ...270...20B> a bibo:Article;
ads:bibcode "1983ApJ...270...20B";
adsbib:author _9:nbachall, _9:rsoneira;
dc:title "The spatial correlation function of rich clusters of galaxies";
dct:subject aakeys:AngularCorrelation, aakeys:MethodsDataAnalysis;
bibo:abstract "A series of objective statistical estimators is applied to directly study the three-dimensional distributi....";
bibo:cites <http://adsset.cfa.harvard.edu:8080/1953ApJ...117..134L>,
<http://adsset.cfa.harvard.edu:8080/1957AJ.....62..248S>,
<http://adsset.cfa.harvard.edu:8080/1958ApJS....3..211A>,
.....;
bibo:pageEnd "38";
bibo:pageStart "20";
bibo:volume "270".

Curation by Humans or Agents!



FOAF



Provenance, Authoring and Versioning



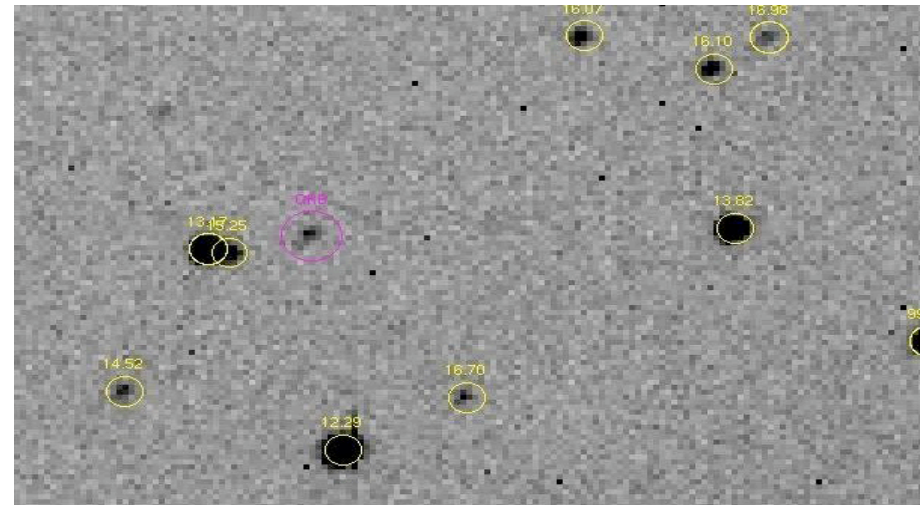
Reification

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
adsbase: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:classIs 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;
adsobject:classIs gcvs:EclipsingBinary
] .

....

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 compatability (data.gov)
- metadata to describe aggregates(ORE)
- interlinking as part of peer-review process?

HELP!

Philosophy: IETF over W3C. Quick and Dirty. Convention over Configuration. Agile. Software.

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.