Astronomy 3.0 Style

Alberto Accomazzi
NASA Astrophysics Data System

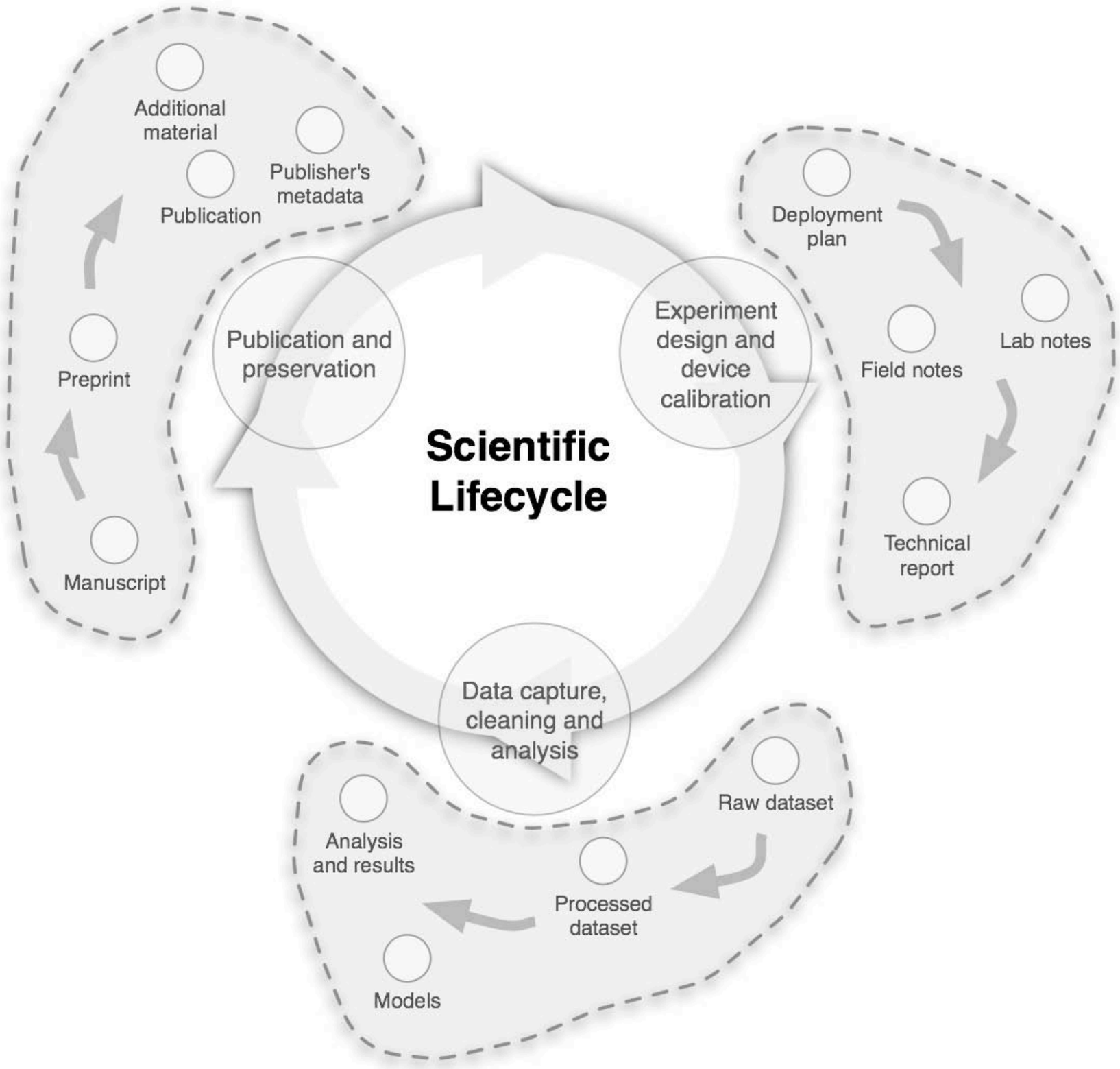
17 Feb 2010
Library and Information Services VI, Pune, India

Overview

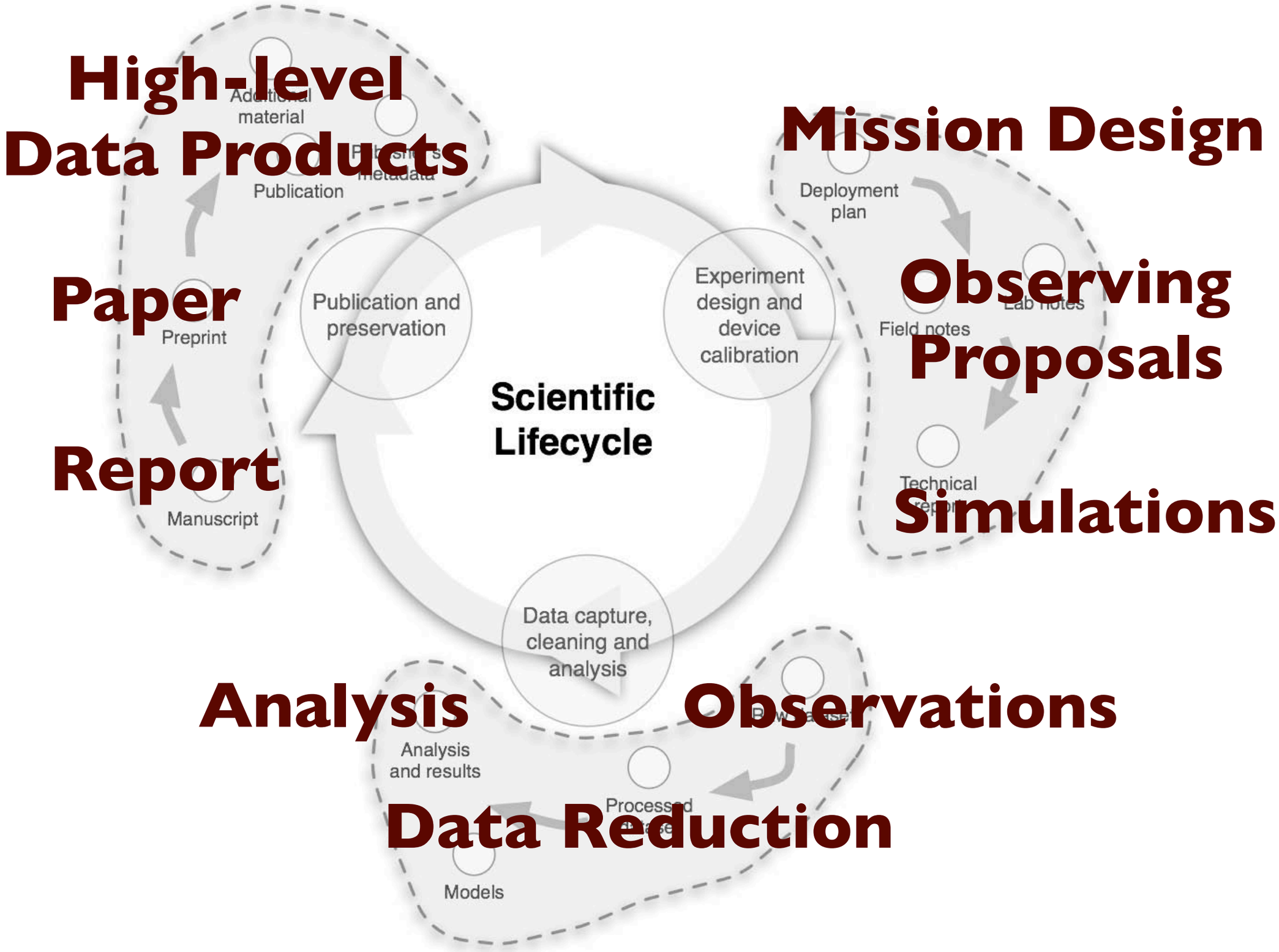
- The Scientific Lifecycle model
- Resource metadata in Astronomy
- What is Web 3.0 and how it can help
- A proposal for interlinking Astronomy resources
- Ongoing work and future plans

The Big Picture

- Scientific research requires repeatability
- The lifecycle of a research project should be documented by capturing all artifacts
- Data, Processes, Results need to be properly described, accessible, and linked together
- Provenance information should be attached to metadata throughout the process



Pepe et al, JASIST (2009)



Pepe et al, JASIST (2009)

Why do we care?

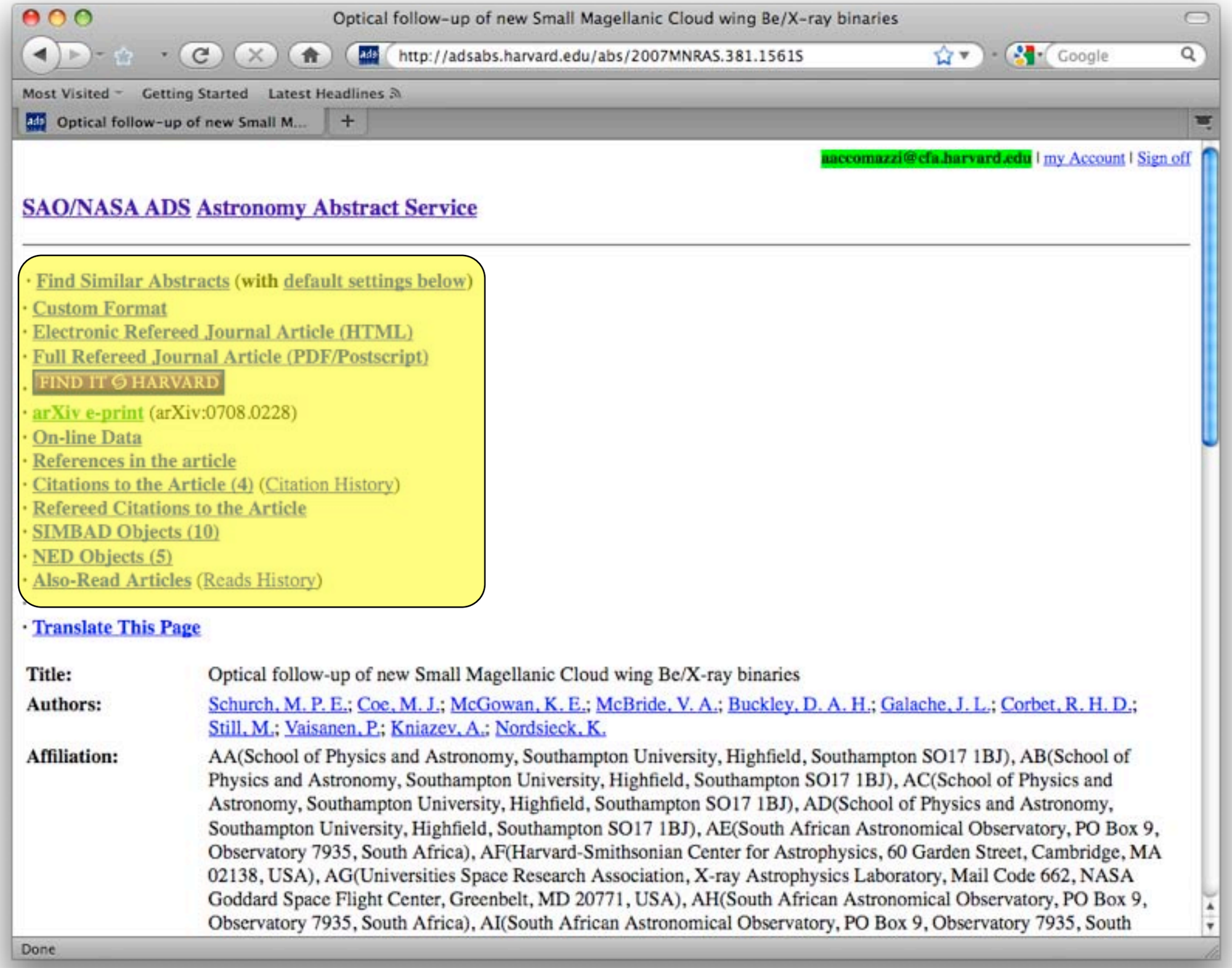
- attribution: links allow one to follow the evolution of research activities
- aggregation: links allow one to find related resources, ancillary data, versioning
- preservation: knowledge about the origin of metadata as important as the data
- discovery: the presence (or absence) of links carries semantic meaning

The case of Astronomy

- Bibliographic Data & Metadata: Publishers, ADS
- Observational Data & Metadata: Archives, Observatories, Surveys
- Object Metadata: NED, SIMBAD, Vizier
- VO to “tie things together”

The Web of Astronomy Links

- Established very early in the game thanks to agreement on identifiers (pre-web!)
- Dense, curated by individual projects on behalf of the community
- Links across domains (papers, data, objects)
- Provides many of the connections required by the science life-cycle model


Optical follow-up of new Small Magellanic Cloud wing Be/X-ray binaries
http://adsabs.harvard.edu/abs/2007MNRAS.381.1561S
Google
Most Visited
Getting Started
Latest Headlines
Optical follow-up of new Small M...
aaccomazzi@cfa.harvard.edu | my Account | Sign off
SAO/NASA ADS Astronomy Abstract Service
· Find Similar Abstracts (with default settings below)
· Custom Format
· Electronic Refereed Journal Article (HTML)
· Full Refereed Journal Article (PDF/Postscript)
· FIND IT @ HARVARD
· arXiv e-print (arXiv:0708.0228)
· On-line Data
· References in the article
· Citations to the Article (4) (Citation History)
· Refereed Citations to the Article
· SIMBAD Objects (10)
· NED Objects (5)
· Also-Read Articles (Reads History)
· Translate This Page
Title:
Optical follow-up of new Small Magellanic Cloud wing Be/X-ray binaries
Authors:
Schurch, M. P. E.; Coe, M. J.; McGowan, K. E.; McBride, V. A.; Buckley, D. A. H.; Galache, J. L.; Corbet, R. H. D.; Still, M.; Vaisanen, P.; Kniazev, A.; Nordsieck, K.
Affiliation:
AA(School of Physics and Astronomy, Southampton University, Highfield, Southampton SO17 1BJ), AB(School of Physics and Astronomy, Southampton University, Highfield, Southampton SO17 1BJ), AC(School of Physics and Astronomy, Southampton University, Highfield, Southampton SO17 1BJ), AD(School of Physics and Astronomy, Southampton University, Highfield, Southampton SO17 1BJ), AE(South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa), AF(Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA), AG(Universities Space Research Association, X-ray Astrophysics Laboratory, Mail Code 662, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA), AH(South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa), AI(South African Astronomical Observatory, PO Box 9, Observatory 7935, South
Done

Your NED Search Results

http://nedwww.ipac.caltech.edu/cgi-bin/nph-objsearch?objr

Recently Bookmarked · DChart -- Ocean, Cli... · Airfare Deals - Chea... · YouTube - Google S... · SKUA project · BaseX - Download · Pubget: the search e...

Your NED Search Results

Data Related Directly to Object Names	Site/Service
Query SIMBAD by primary NED object name -- MESSIER 031	SIMBAD (CDS, Strasbourg, France)
Revised New General Catalogue -- NGC 0224	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
Uppsala General Catalog -- UGC 00454	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
Original Zwicky Catalog -- CGCG 535-017	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
Query UZC Spectral Archive (60 arcsec search radius)	Updated Zwicky Catalog Data (Harvard/SAO)
Morphological Catalog of Galaxies -- MCG +07-02-016	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
The Third Bologna Survey -- B3 0040+409	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
2MASS Extended Source Images (JHKs) -- 2MASX J00424433+4116074	NASA/IPAC Infrared Science Archive (IRSA)
IRAS Point Source Catalog -- IRAS 00400+4059	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
IRAS Faint Source Catalog -- IRAS F00400+4059	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
Catalogue of Principal Galaxies -- PGC 002557	VizieR Catalog Query (U.S. mirror, CfA/Harvard)
Retrieve mean data from LEDA -- PGC 002557	The Lyon/Meudon Extragalactic Database (LEDA)
Query GALEX (NUV/FUV) Mission Archive (6' search radius) -- MESSIER 031	GALEX Mission Data Archive at MAST
General Archive Resources -- All queries centered at 00h42m44.3s, +41d16m09s (J2000)	**Site/Service**
Query Optical and UV Mission Archives (Default search radius)	Multimission Archive at STScI (MAST)
Query High Energy Mission Archives (Default search radius)	HEASARC (NASA/GSFC)
Explore resources with DataScope (15' search radius)	HEASARC (NASA/GSFC)
Visualize Coverage Map with IMPReSS Size: 1 deg	Astrophysics Data Facility (NASA/GSFC)
Retrieve 2MASS Atlas Images Band(s): Ks Size: 2'	NASA/IPAC Infrared Science Archive (IRSA)
Retrieve IRAS ISSA Images Band(s): 60um Size: 30'	NASA/IPAC Infrared Science Archive (IRSA)
1-D Coadd of IRAS Scans (ADDSCAN/SCANPI)	NASA/IPAC Infrared Science Archive (IRSA)
Retrieve NVSS Image Size: 15' ⊙ Contours (PS) ○ JPEG ○ FITS File	NRAO/VLA Sky Survey (NVSS)

Done

Some numbers for ADS

- citations: 40M (across databases)
- readership: 18M (90-day window)
- fulltext: 5M (journals, arXiv, ADS)
- astronomical objects: 240K (SIMBAD + NED)
- data products: 130K
- bibliographic groups: 200K

Bibliographic groups

- CfA
- ESO
- ROSAT
- HST
- IUE
- CFHT
- Chandra
- ISO
- PhysEd
- XMM
- USNO
- NRAO
- Gemini
- Keck
- Spitzer

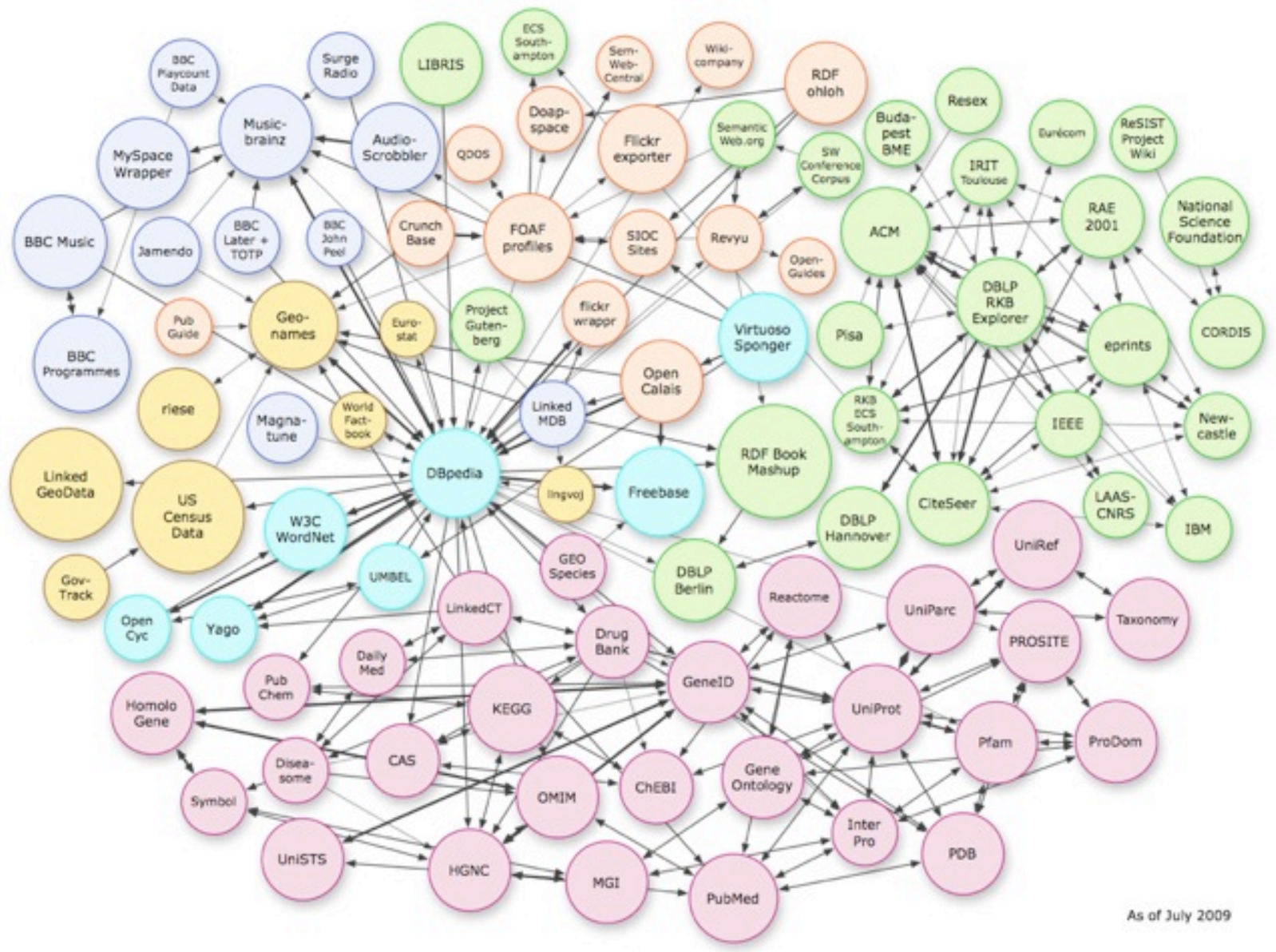
Web 3.0

- The "Web of Data" or "Semantic Web"
- Describes resources in a formal way
- Machine-readable content throughout
- Based on Linked Open Data principles
- "Read, Write, Execute"

Linked Data principles

- 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*

Linked Data Cloud

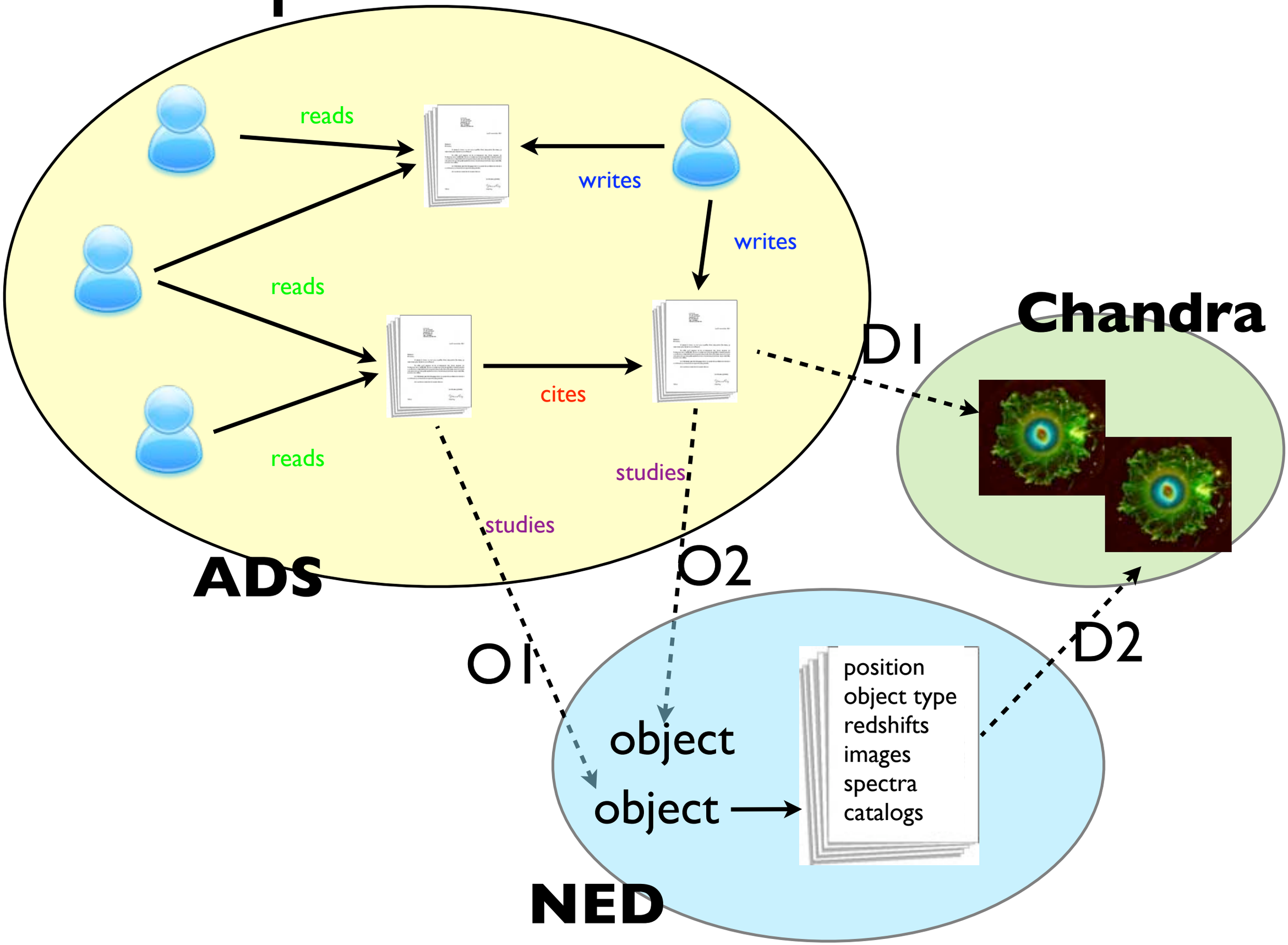


as of jan 2010: 142M links, 13B triples

Links to Datasets

- have been based on reciprocal links between URLs curated by collaborators
- do not identify resources as unique URIs
- do not make use standard vocabularies to describe data or the types of links between them
- are not actionable by applications

Example: ADS/NED/Chandra



How to get there

- grow the astronomy data cloud from the bottom-up
- identify our resources (data), name them uniquely, expose their metadata
- expose relationships between resources (links)
- what should be included: observing proposals, observational metadata, instrumental metadata, papers, catalogs, objects

Publication Metadata

- Authors
- Keywords
- Affiliations
- Citations
- Links to data, objects
- Readership

Observational Metadata

- Unique Dataset Identifiers
- Position, footprint and date
- Objects observed
- Wavelengths
- Instruments
- People, Proposals, Collaborations

A VAO Proposal

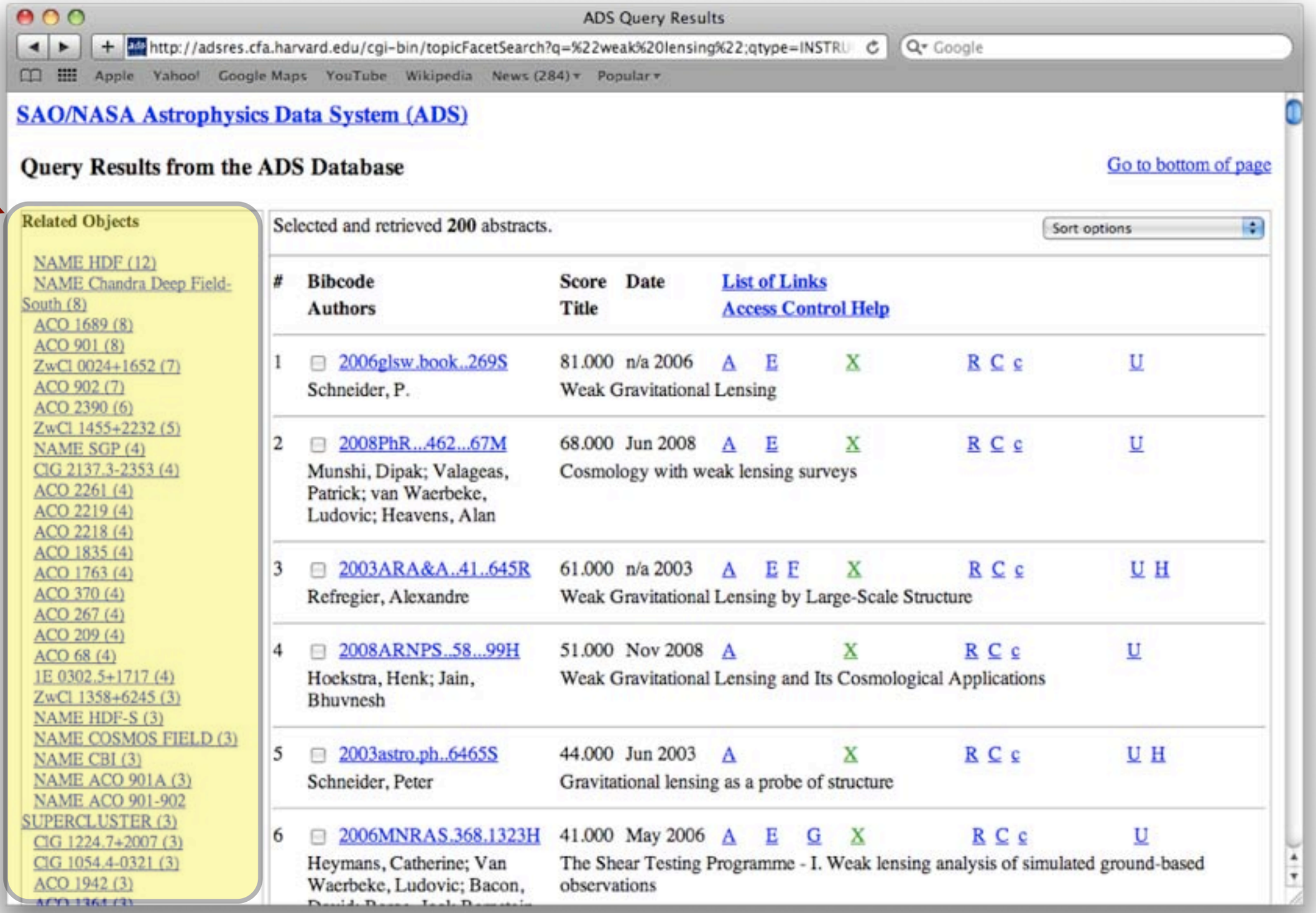
- Identify, collect and expose metadata for Datasets (Chandra, MAST, TSC, etc.)
- Create typed links between Data products, Bibliographic and Object databases
- In the process, create knowledge base about Instruments and Services
- Incorporate all of the above in a "metadata store," exposing resources using semantics

Sample Applications

- A view of bibliographies with facets based on objects, wavelengths, keywords, etc
- A recommender / alert service for data products, objects
- A UI integrating views of bibliographies, objects, and datasets interacting together
- An application computing metrics on data, objects based on citation and use of papers

Views on Papers

object facets



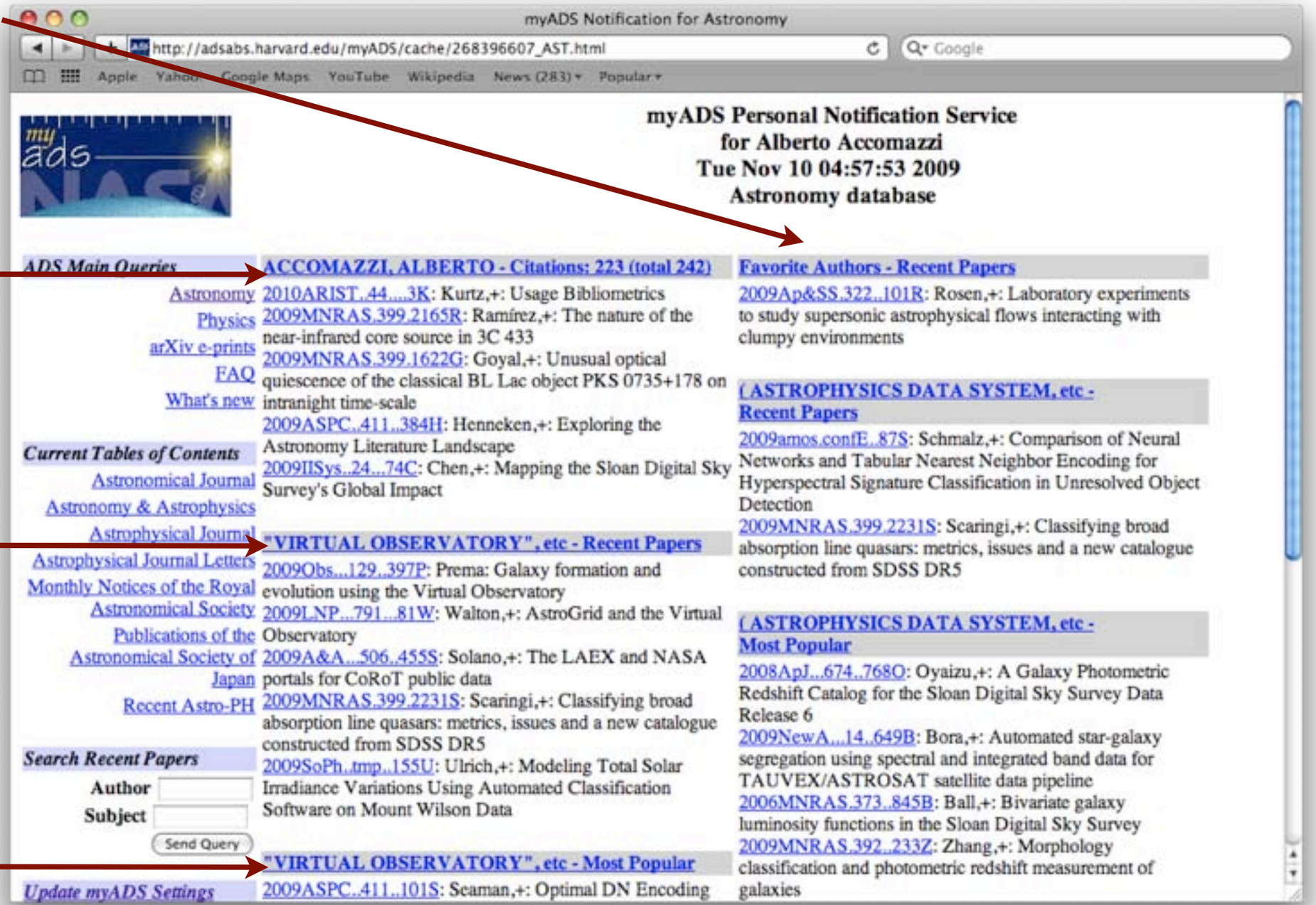
Recommendations?

Collaborations

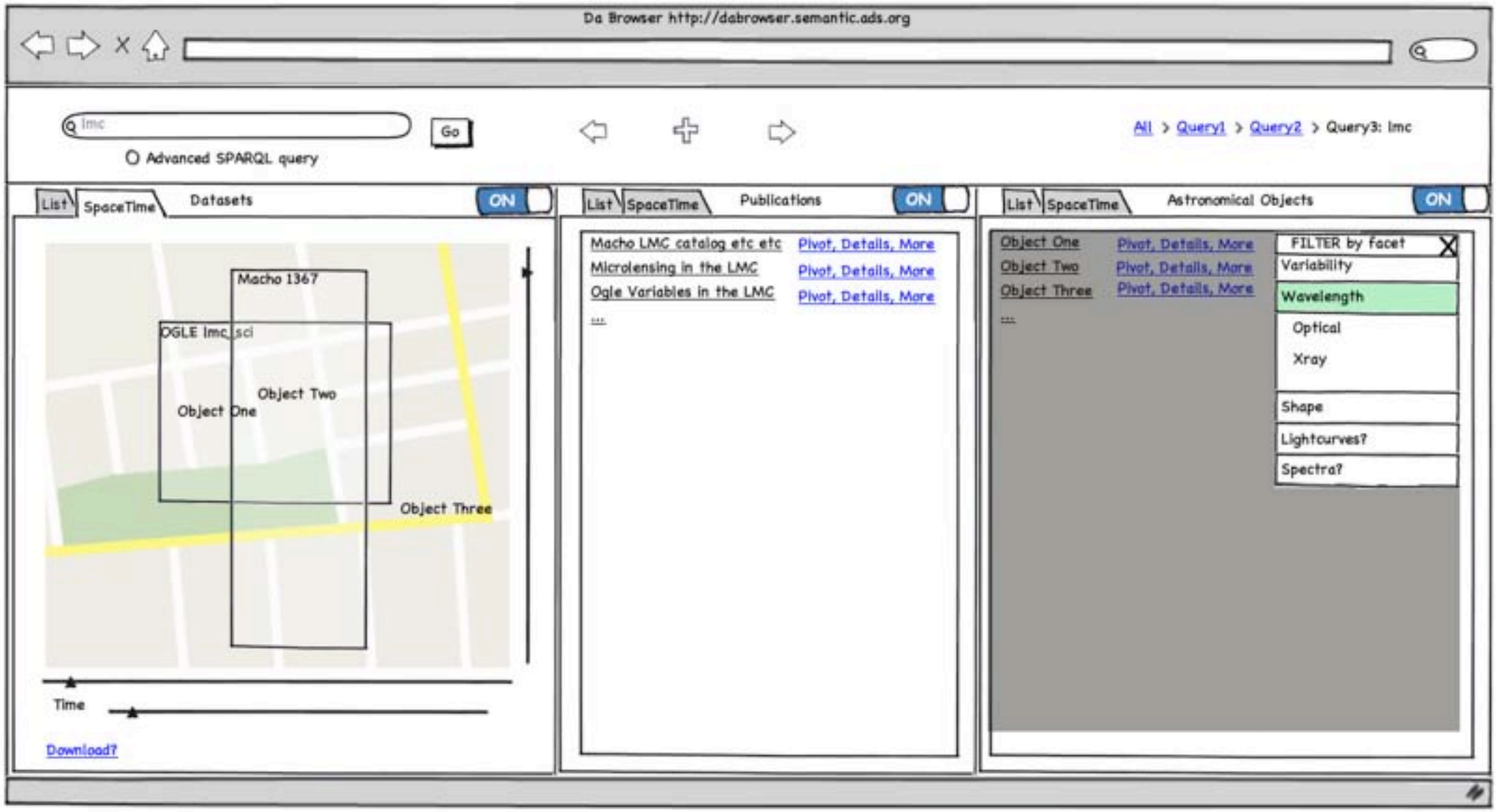
My data

Recent data

Hottest data



APOD Application



Ongoing Efforts

- IVOA / ADEC dataset identifiers
- IVOA data models, vocabularies for keywords and UCDs, ontology of object types
- MSR / WWT work on mashups
- ADS work with VAO, text mining, faceted views

The Role of Libraries

- Astronomy librarians have been at the intersection of archives and journals
- They are experts in metadata curation and search, maintain bibliography-data links
- They have been collecting/generating much of this metadata already
- They understand the need for (and limitations of) metrics

A call to action

- Tell us what bibliographic and observational metadata you are collecting and curating
- Use our upcoming fulltext search on current literature to collect/annotate papers
- Drink the Web 3.0 coolaid
- We have freed our data, now let's free your metadata!