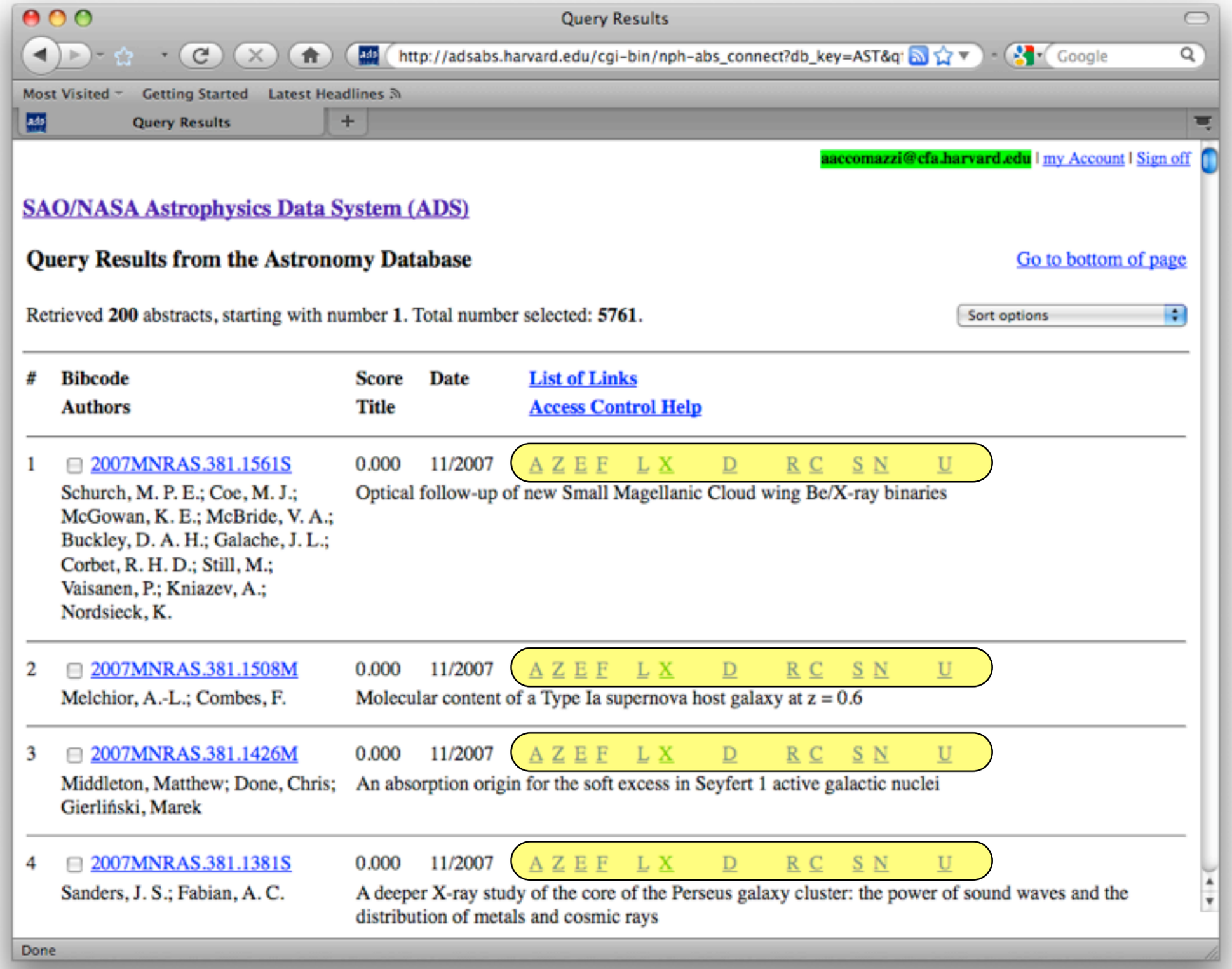
Links in the Astronomy data network

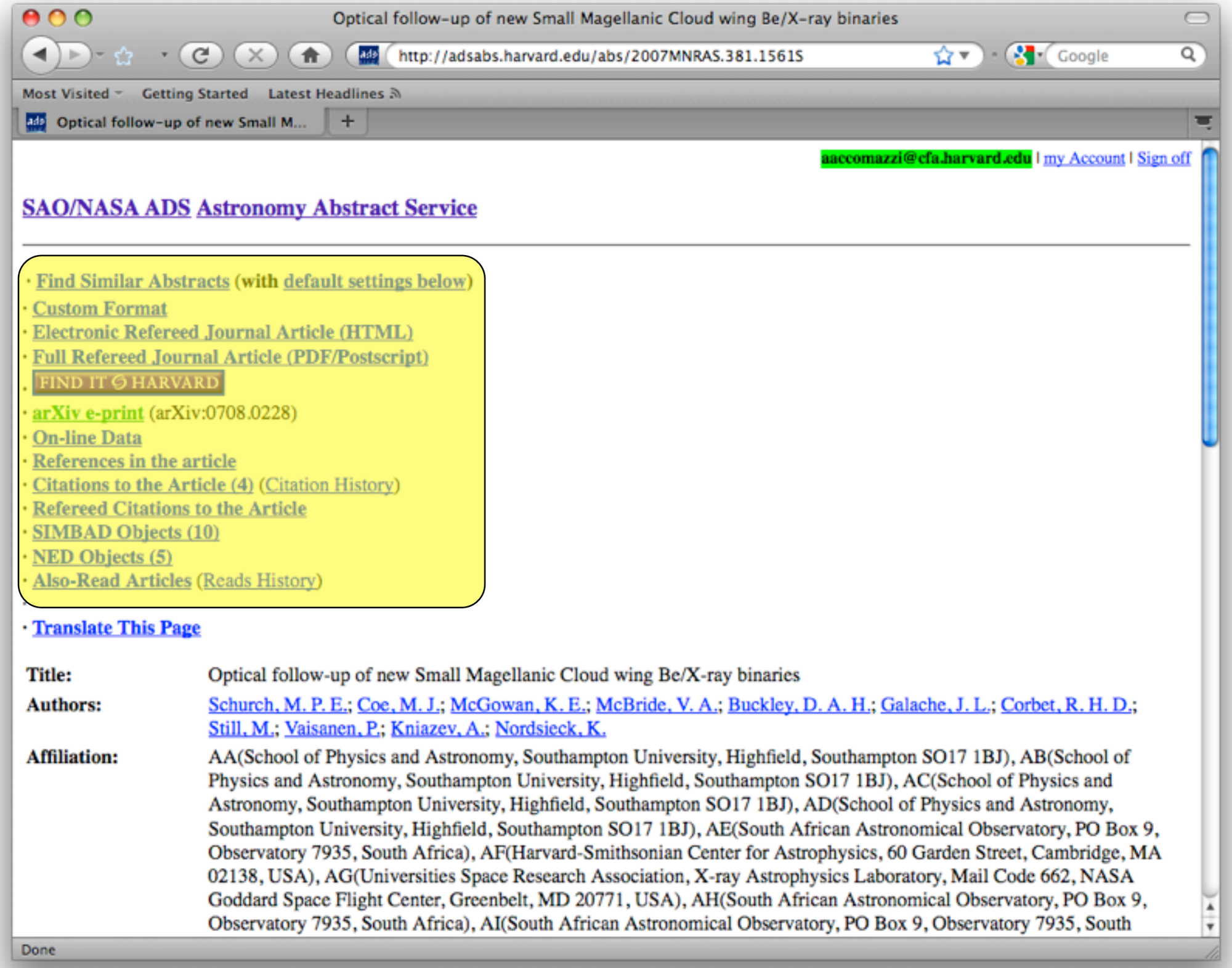
Alberto Accomazzi

The NASA Astrophysics Data System
http://ads.harvard.edu

Bibliographic links

- if papers represent the body of knowledge...
- ...then links are the connective tissue of the body
- as the body evolves, additional connections are created, changing its anatomy
- links provide context, meaning and weight to the papers (e.g. pagerank)


Query Results
http://adsabs.harvard.edu/cgi-bin/nph-abs_connect?db_key=AST&q
Google
Most Visited
Getting Started
Latest Headlines
Query Results
aaccomazzi@cfa.harvard.edu | my Account | Sign off
SAO/NASA Astrophysics Data System (ADS)
Query Results from the Astronomy Database
Go to bottom of page
Retrieved 200 abstracts, starting with number 1. Total number selected: 5761.
Sort options
Bibcode Score Date List of Links
Authors Title Access Control Help
1 2007MNRAS.381.1561S 0.000 11/2007 A Z E F L X D R C S N U
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.
Optical follow-up of new Small Magellanic Cloud wing Be/X-ray binaries
2 2007MNRAS.381.1508M 0.000 11/2007 A Z E F L X D R C S N U
Melchior, A.-L.; Combes, F.
Molecular content of a Type Ia supernova host galaxy at z = 0.6
3 2007MNRAS.381.1426M 0.000 11/2007 A Z E F L X D R C S N U
Middleton, Matthew; Done, Chris; Gierliński, Marek
An absorption origin for the soft excess in Seyfert 1 active galactic nuclei
4 2007MNRAS.381.1381S 0.000 11/2007 A Z E F L X D R C S N U
Sanders, J. S.; Fabian, A. C.
A deeper X-ray study of the core of the Perseus galaxy cluster: the power of sound waves and the distribution of metals and cosmic rays
Done


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

Types of ADS links

- internal (citations, similar articles)
- external (fulltext, data, catalogs, objects)
- computed (citations, usage, openURL)
- curated (data products, associated articles)
- ADS-generated (usage, citations, fulltext)
- contributed (data links, objects)

Some numbers for ADS

- citations: 36M (across databases)
- readership: 15M (90-day window)
- fulltext: 5M (journals, arXiv, ADS)
- astronomical objects: 240K (SIMBAD + NED)
- data products: 130K

Bibliographic groups

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

Thank you for your contributions!

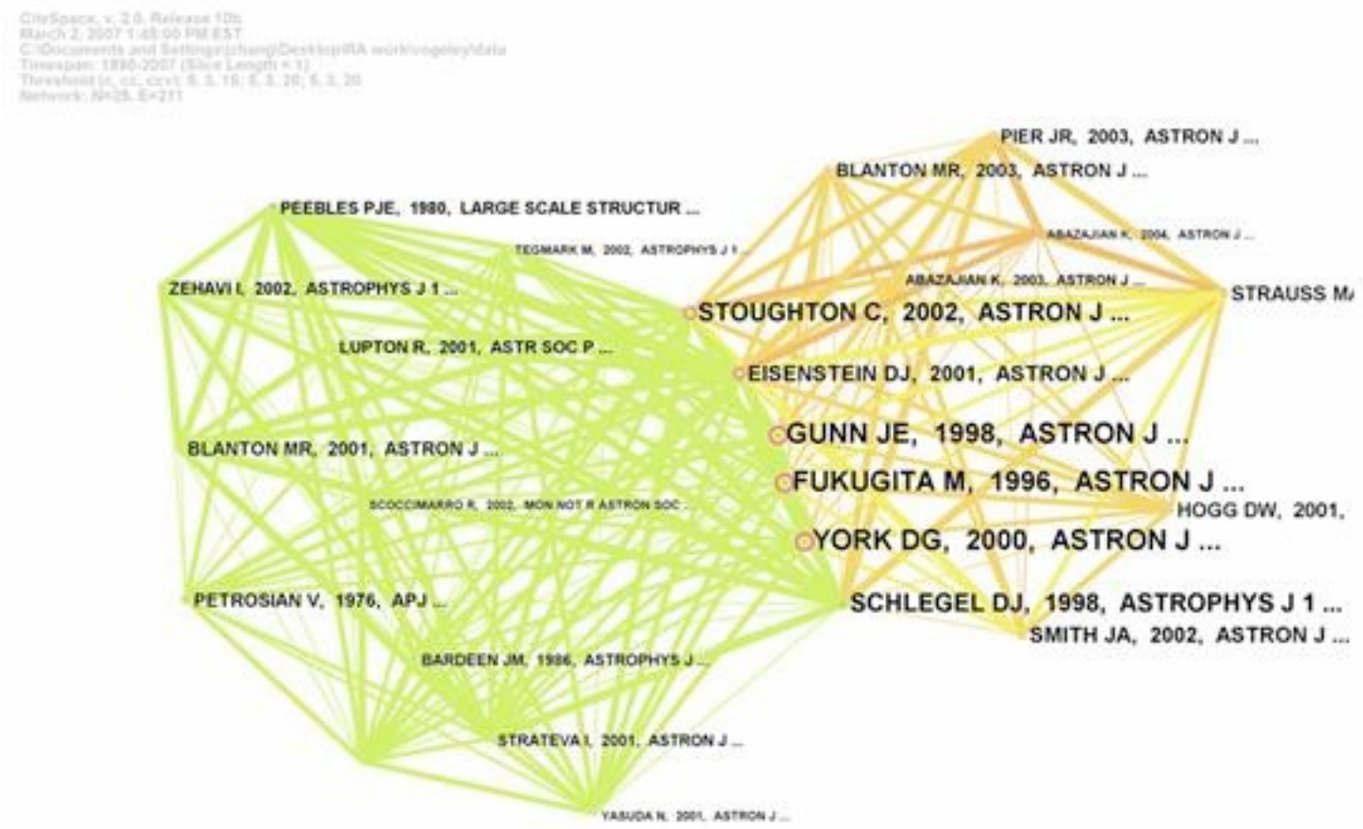
Why do we care?

- attribution: citation links allow one to follow the evolution of science
- aggregation: data links allow one to find resources that are useful in research
- preservation: the value of resources is dependent on their environment
- discovery: the presence (or absence) of links carries semantic meaning

Links in the brave new world

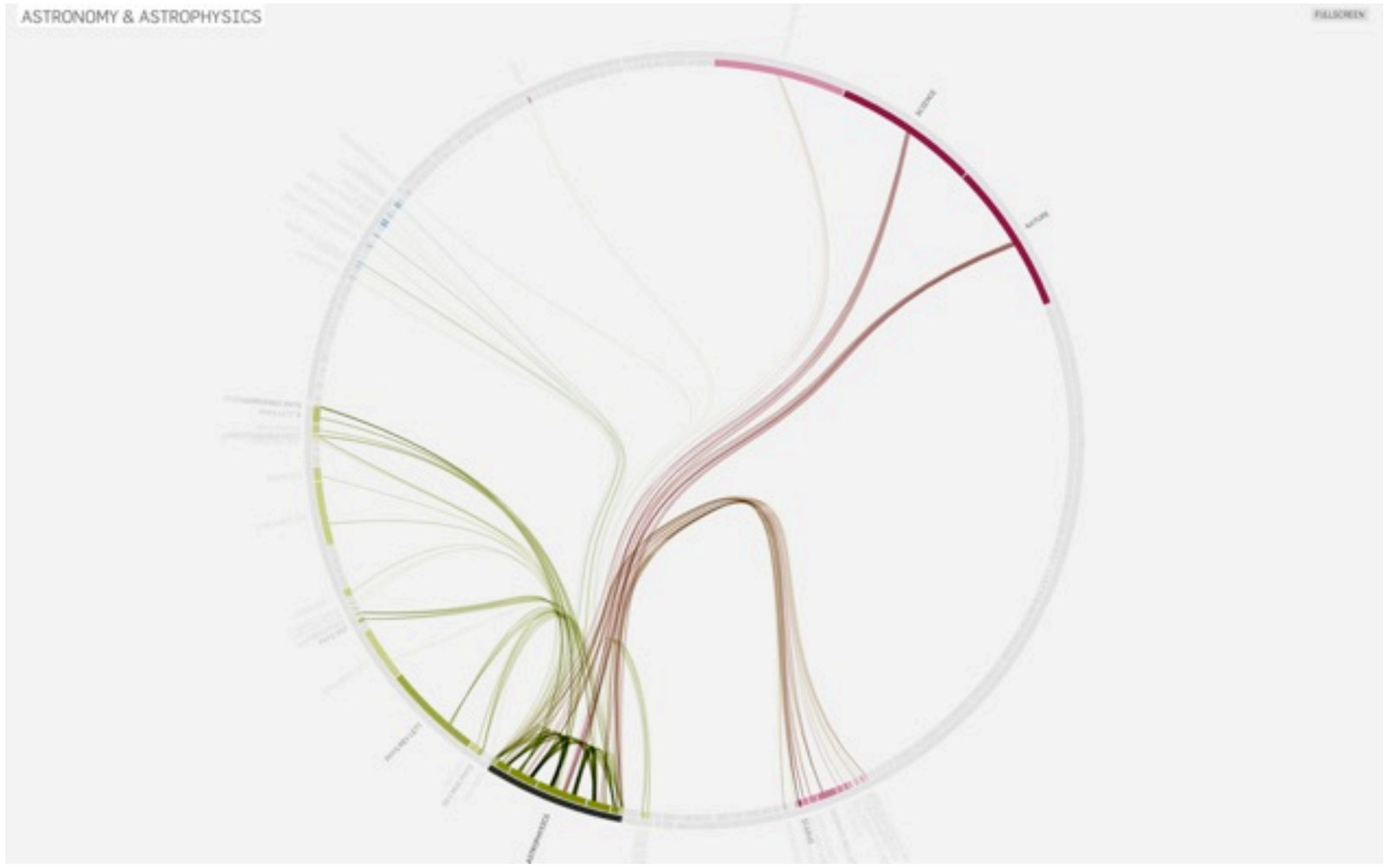
- Graph theory, network analysis tools can model and analyze information flow
- The Semantic Web and Linked Open Data effort broaden the data network
- Object Reuse and Exchange (OAI/ORE) provides framework for describing aggregations of (scholarly) web resources

Citation Networks



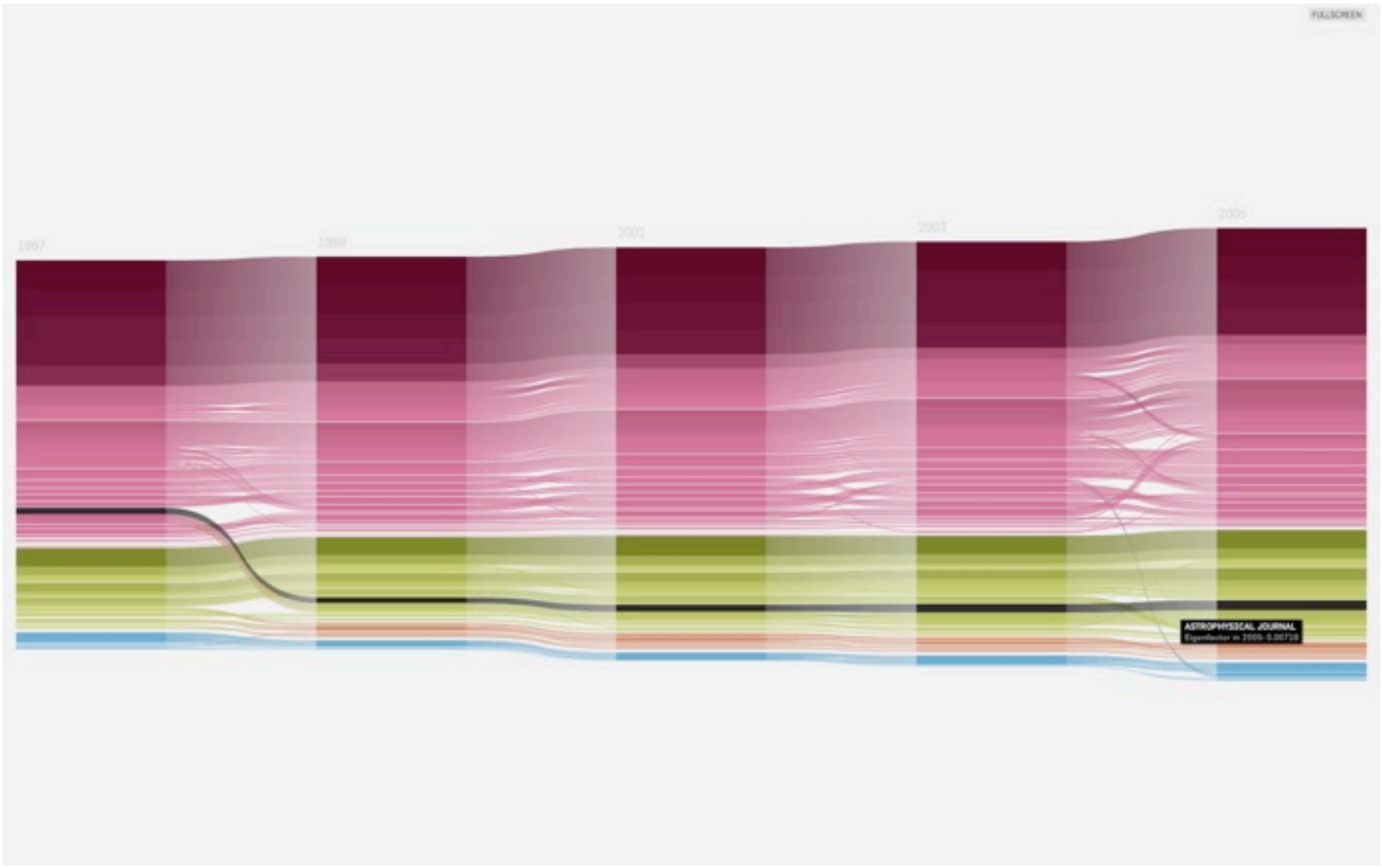
SDSS papers mapped using CiteSpace (Chaomei Chen)

Citation Patterns



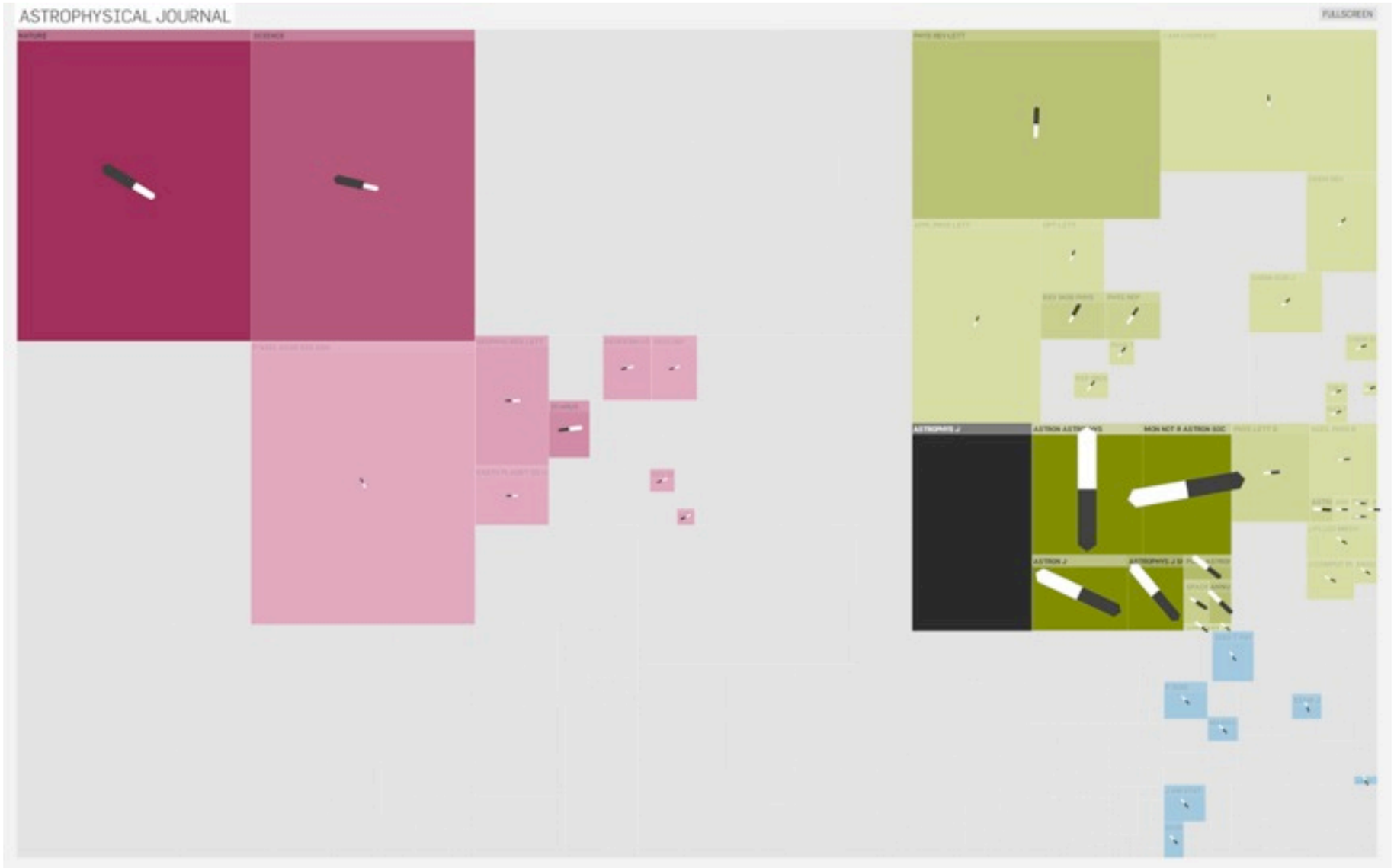
eigenvector.org: Astronomy and Astrophysics journal citations

Change over time



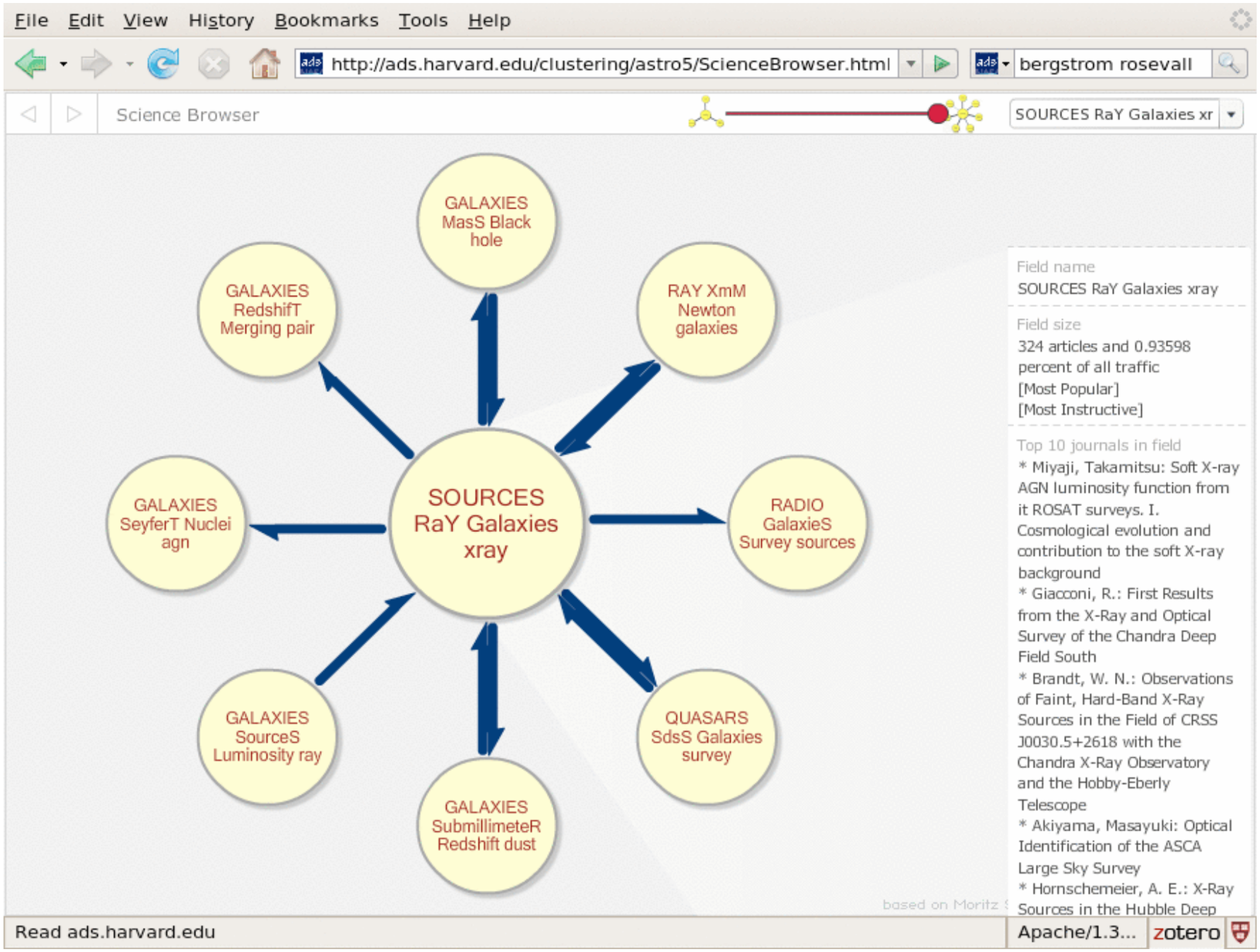
eigenfactor.org: change in impact and clustering

Treemaps



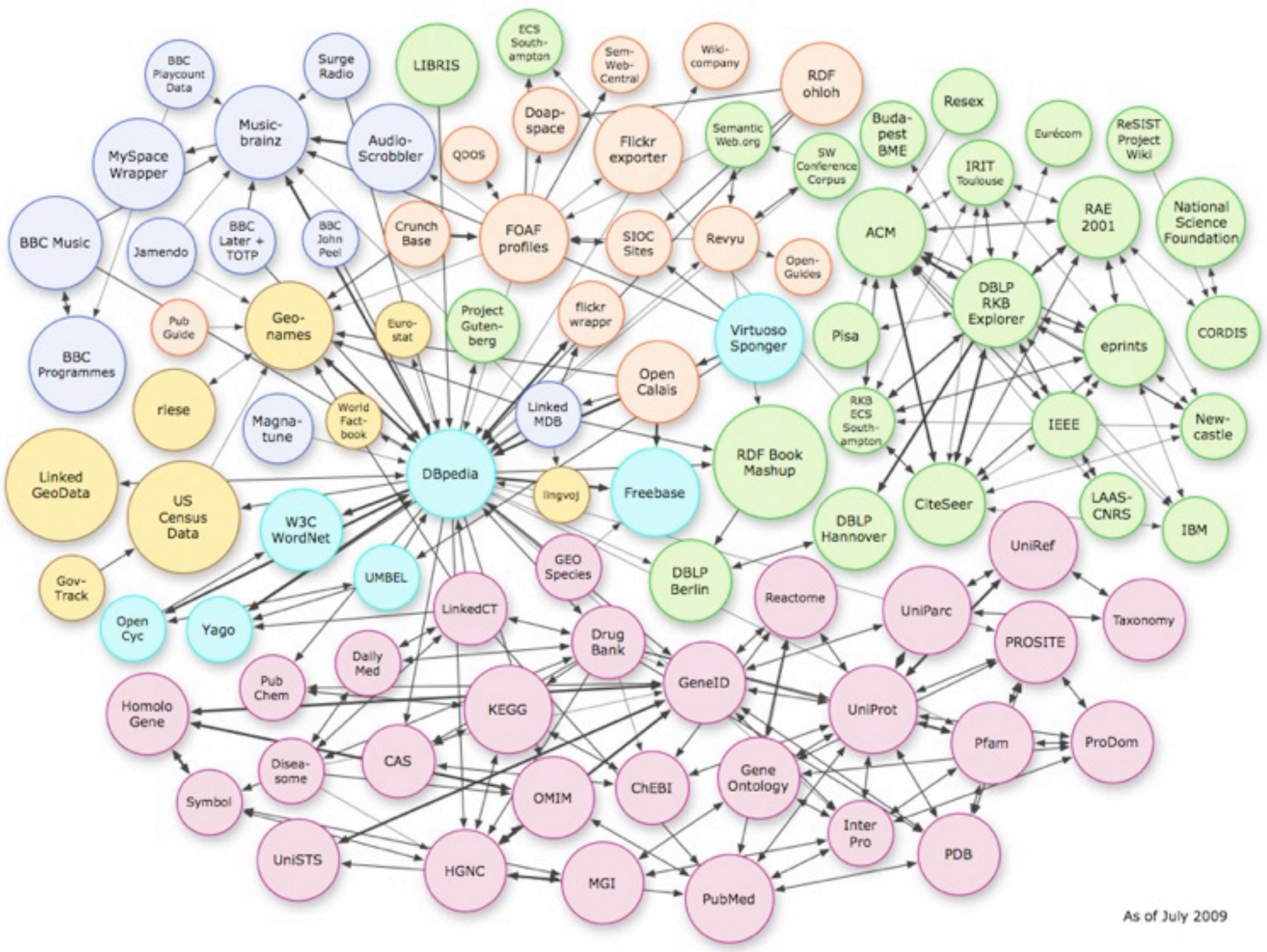
eigenfactor.org: journal citation flow

ADS topic clusters



clusters generated from an ADS citation graph

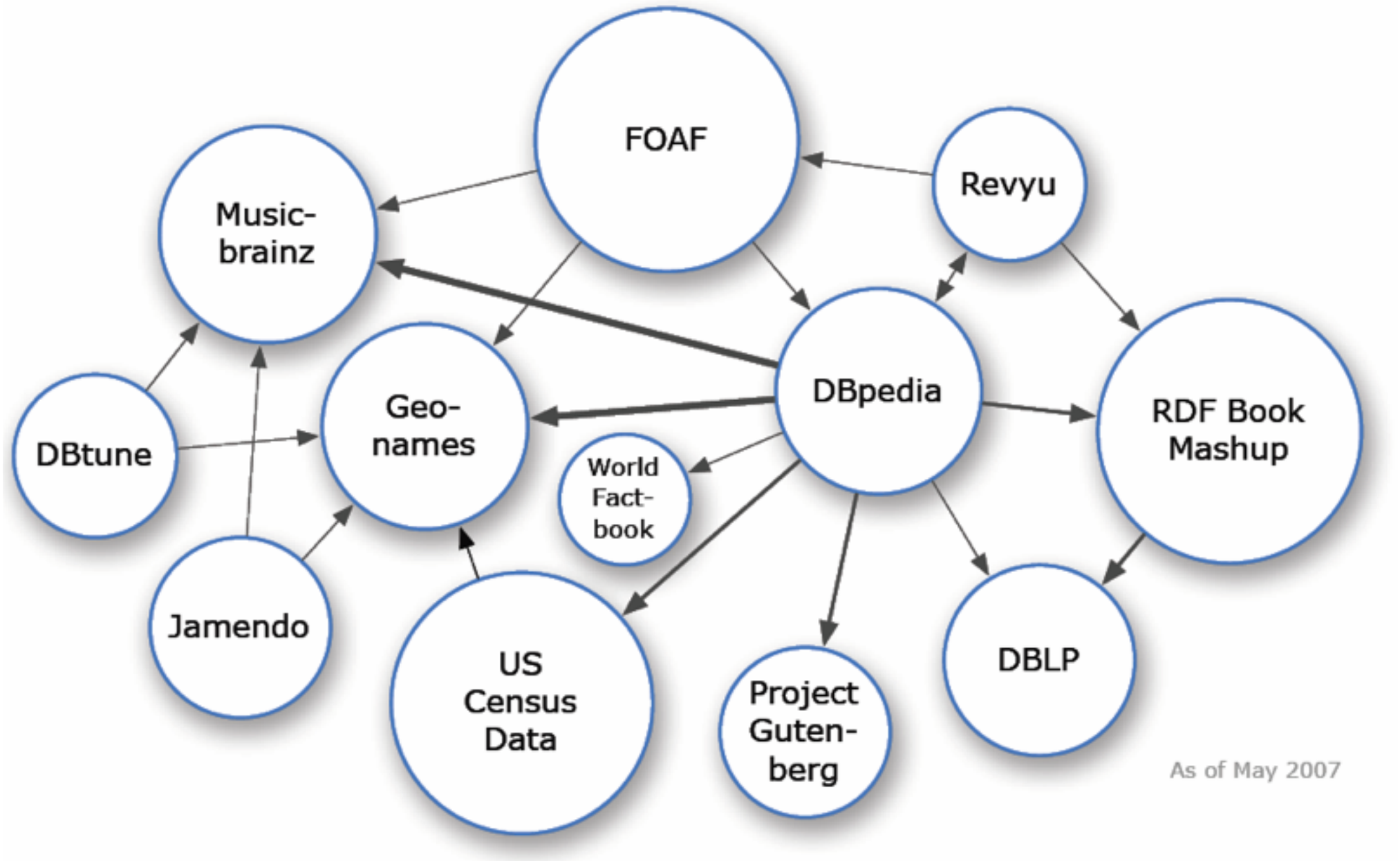
Linked Open Data



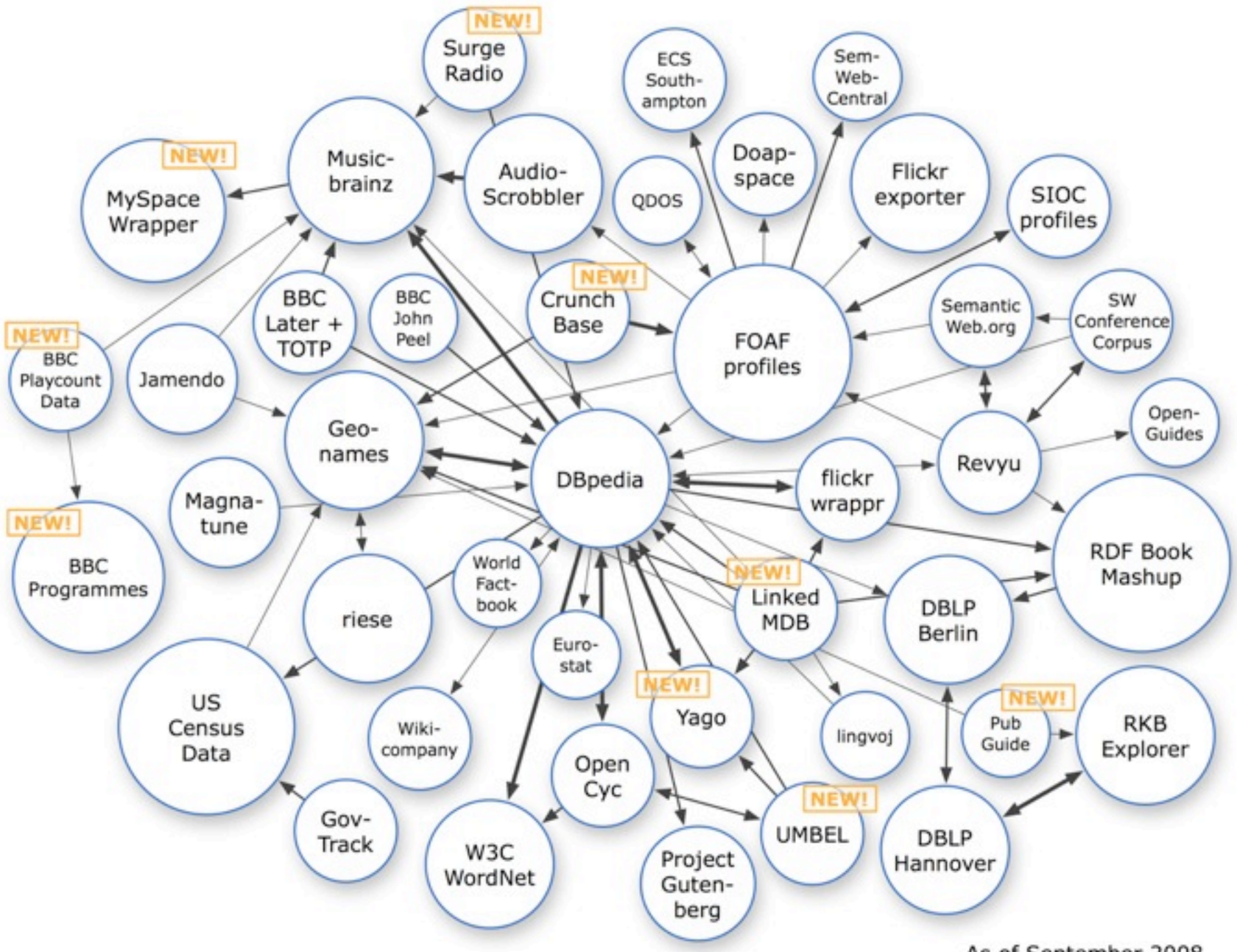
LOD principles

- resources are named via HTTP *URIs*
- metadata is *open* and in a standard format (RDF)
- links between resources are *typed*
- built on the architecture of the web, *no APIs*

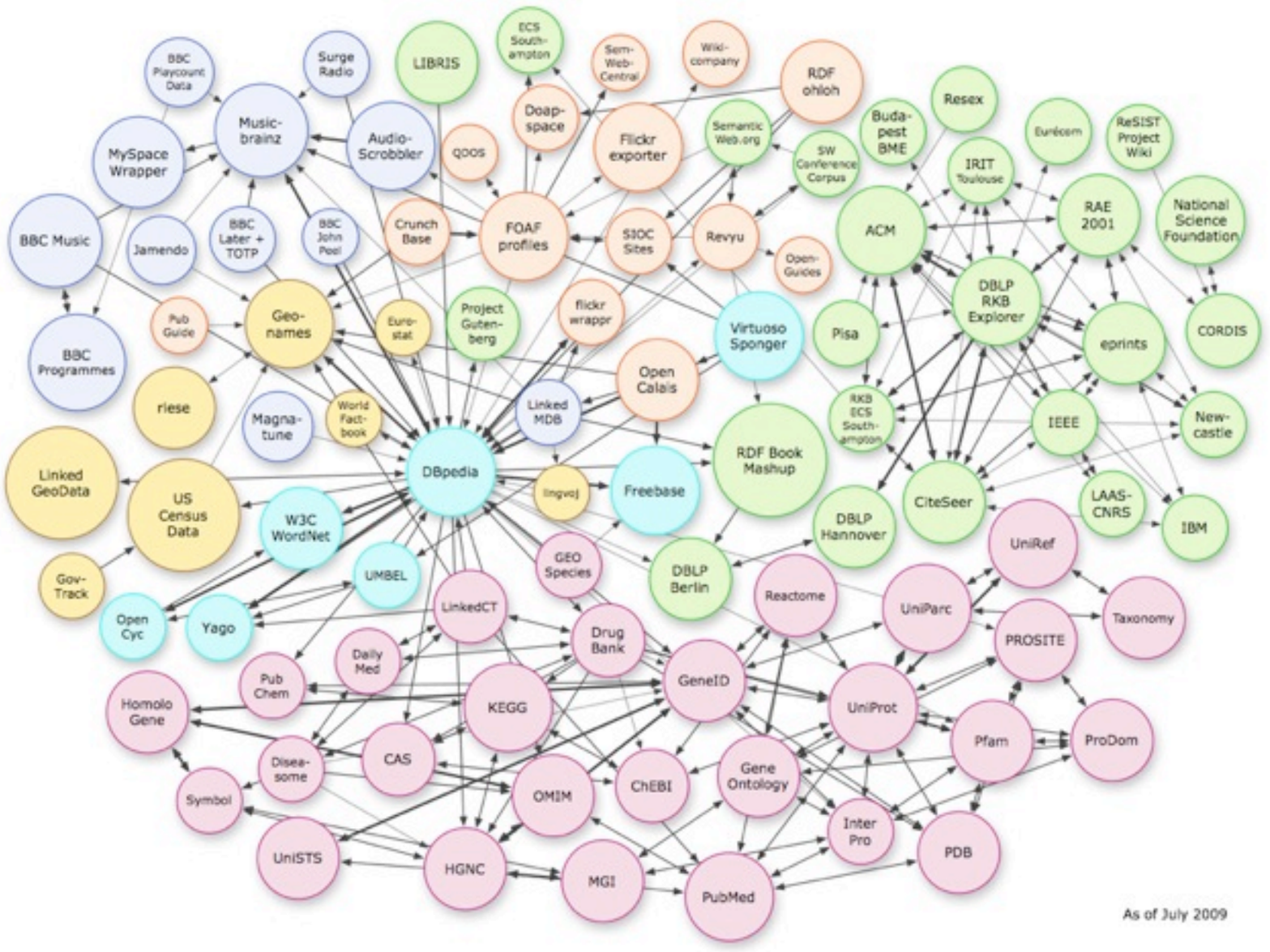
The first LOD cloud



It grows...


NEW!
Surge Radio
ECS South-ampton
Sem-Web-Central
NEW!
MySpace Wrapper
Music-brainz
Audio-Scrobbler
QDOS
Doap-space
Flickr exporter
SIOC profiles
NEW!
BBC Playcount Data
BBC Later + TOTP
BBC John Peel
NEW!
Crunch Base
FOAF profiles
Semantic Web.org
SW Conference Corpus
Jamendo
Geo-names
DBpedia
Revyu
Open-Guides
Magna-tune
flickr wrappr
NEW!
BBC Programmes
World Fact-book
RDF Book Mashup
riese
NEW!
Linked MDB
DBLP Berlin
Euro-stat
US Census Data
NEW!
Yago
NEW!
Pub Guide
RKB Explorer
Wiki-company
lingvoj
Open Cyc
Gov-Track
NEW!
UMBEL
W3C WordNet
Project Guten-berg
DBLP Hannover
As of September 2008

And grows...

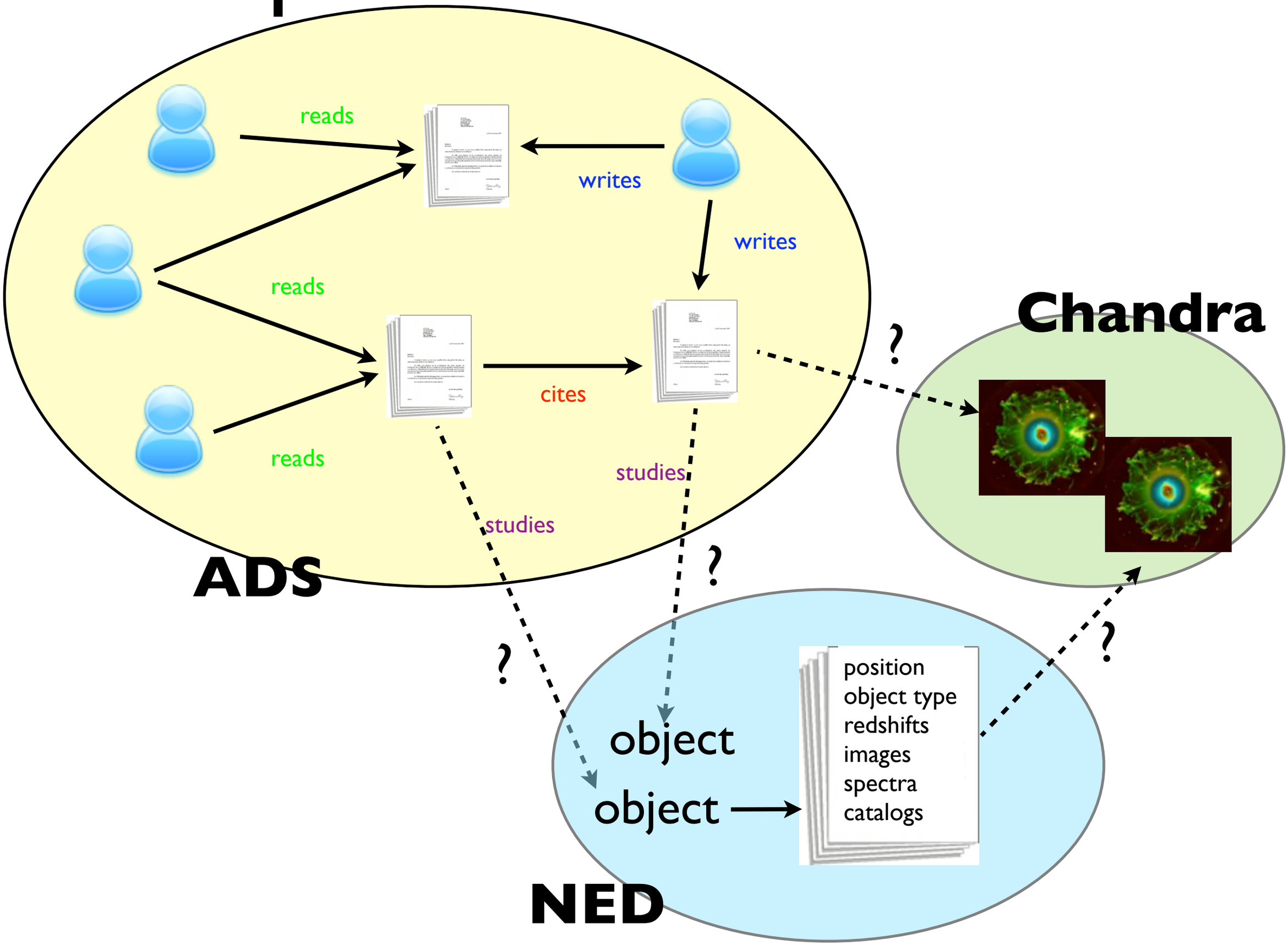


As of may 2009: 142M links, 4.2B triples

The Astronomical cloud

- has been based on reciprocal links between URLs curated by collaborators
- does not describe resources or the types of links between them
- does not use standard vocabularies to describe things
- is not actionable by applications

Example: ADS/NED/Chandra



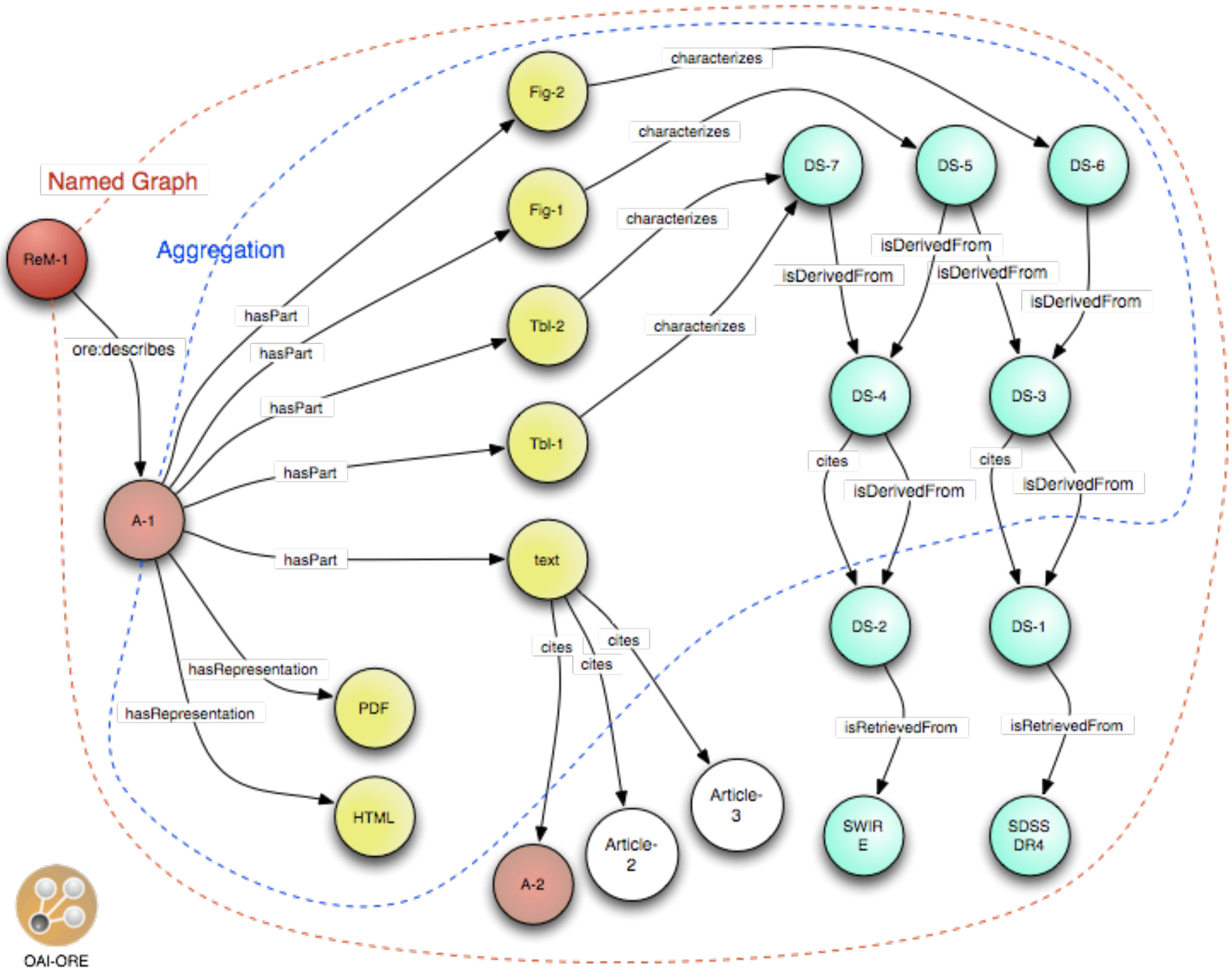
How to get there

- the astronomy data cloud should be "grown" 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

Object Reuse & Exchange

- defines standards for the description and exchange of web resources
- is based upon LOD principles
- used to describe the scholarly "value chain"
- publications are no longer independent entities but rather complex objects
- preservation "done right" for e-science

OAI/ORE in action



Ongoing projects

- ADS to exploit user/paper/keyword networks for recommendations, faceting
- ADS/VAO/MSR proposal using Linked Data approach to power semantically aware apps
- JHU Libraries project on curating publications
- Datanet Data Conservancy project (JHU)

Conclusions

- in a densely interconnected world, curating links becomes essential
- expect to see a number of applications exploiting the network of links
- technology based on RDF, LOD, OAI/ORE
- be part of the graph, don't be the end leaf