New ADS Developments

Alberto Accomazzi
Michael Kurtz

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

Highlights of past year

- management and infrastructure

- data holdings

- collaborations

- future plans

Management & Infrastructure

- ADS Ranked #1 in NASA's astrophysics archives review

- finally about to be fully staffed @ 7.5FTE

- now implementing a sound disaster recovery plan

- working plan for system restructuring

Data Holdings

- integrated 4.5M records from CrossRef

- bibliographic metadata coverage up by 25% (to 7.7M)

- citations up by 28% to 35M

- ingested 180K pages of historical fulltext (3.8M total)

Collaborations

- CERN/CDS: ADS to adopt and extend Invenio as DL software platform

- SPIRES/INSPIRE: e-print/DOI mappings, HEPnames, curation workflow

- CrossRef: registration of historical records

- VAO: biblio-enabled "data" search

Projects

- document classification

- enhanced fulltext search

- contextual views

- semantic search

- "data" search

Document Classification

- document classification and categorization

- makes use of different clustering techniques making use of citations, keywords

- relates to document as a whole rather than its parts
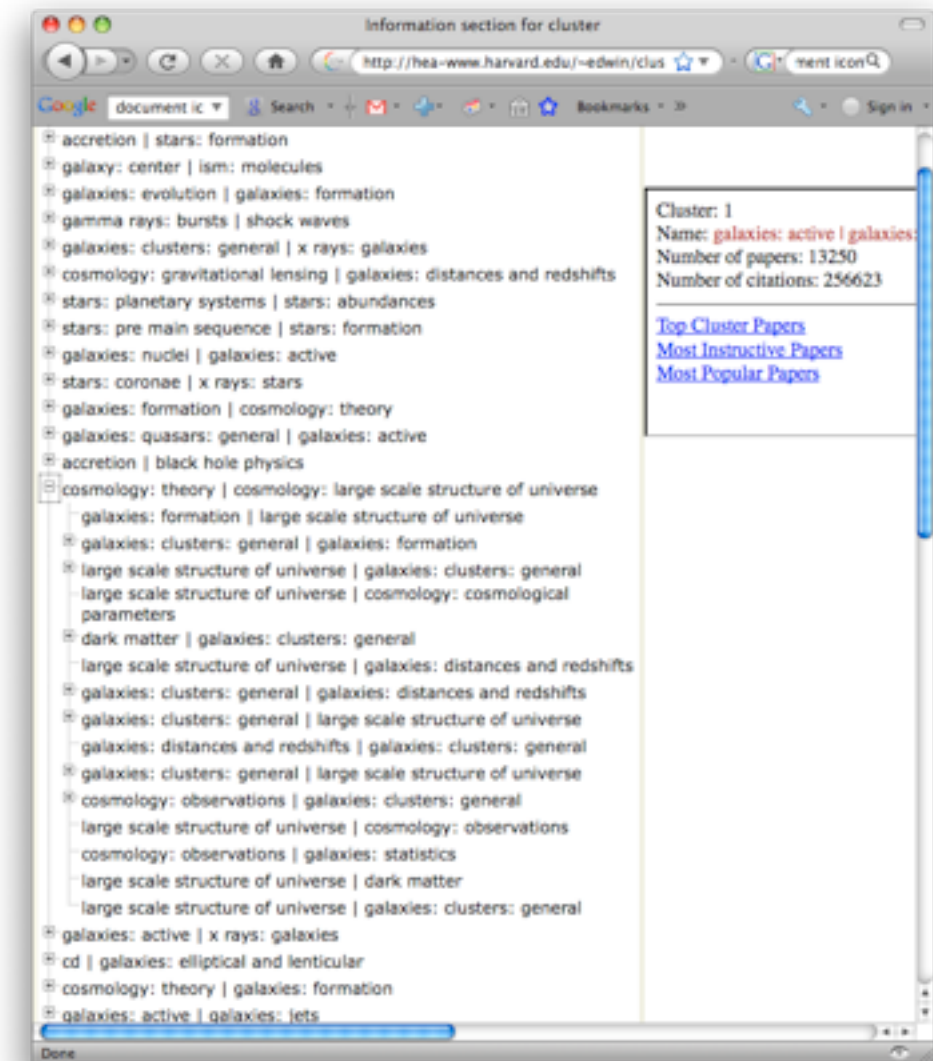
Fulltext search

- ADS to become "dark archive" of current astronomy fulltext

- ADS fulltext search will enable finding current and historical content

- text mining will provide tools to formalize and extract information from fulltext documents

- text summarization and entity extraction by Murphy & Currant, SIMBAD

Contextual views

- ADS has experimented with applets for interactive topic visualization

- CDS has used Kohonen self-organizing topic maps

- PaperScope displays a paper in citation network
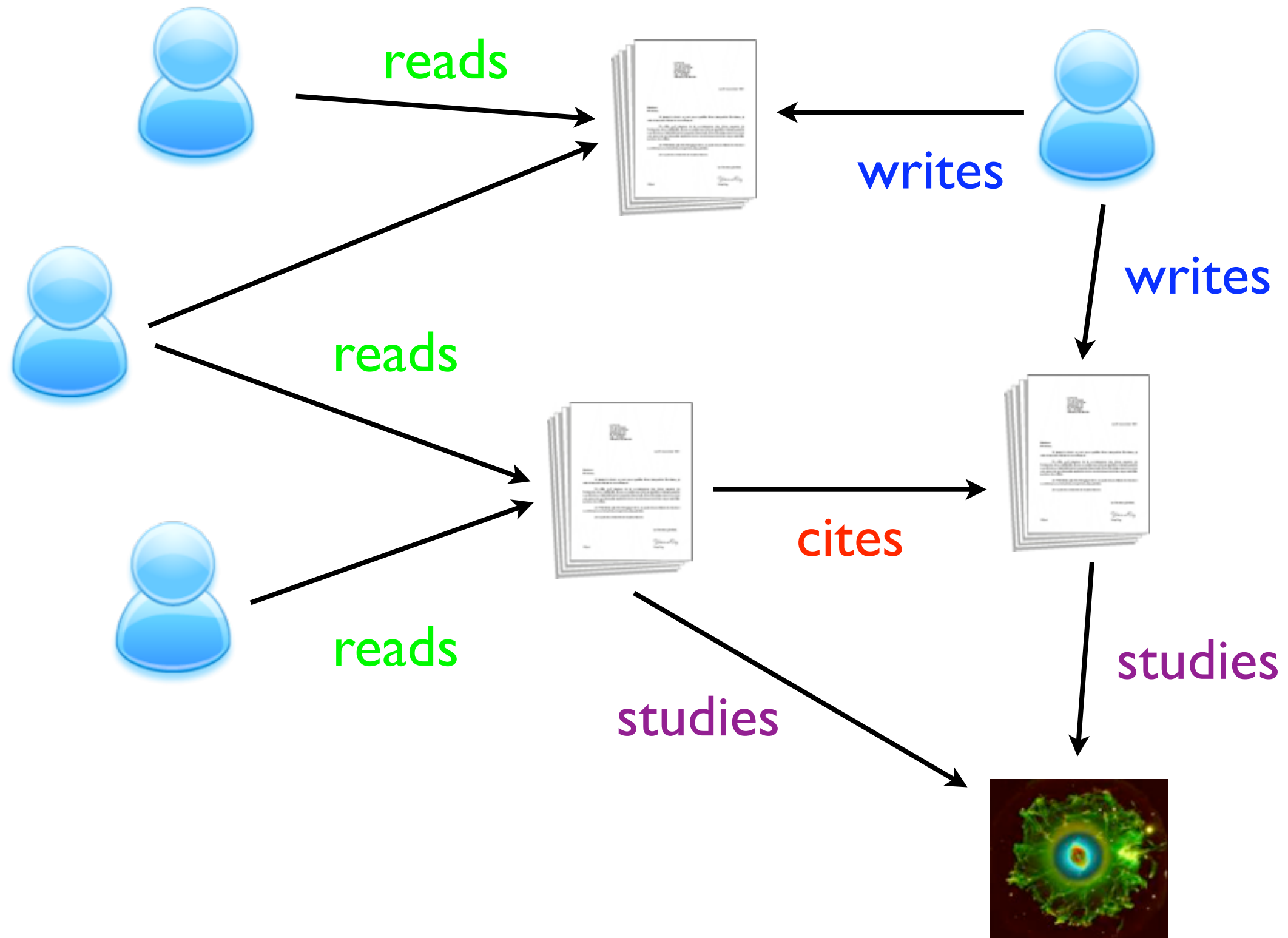
- no "killer app" yet
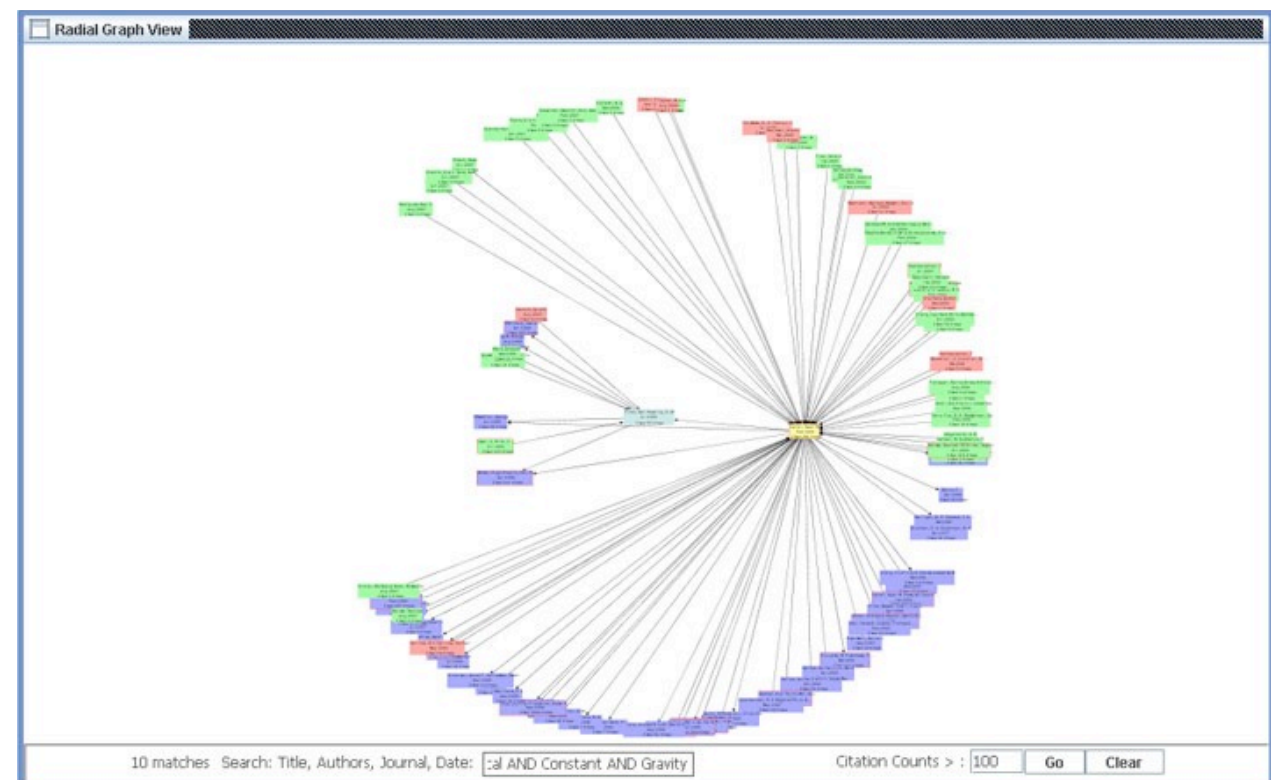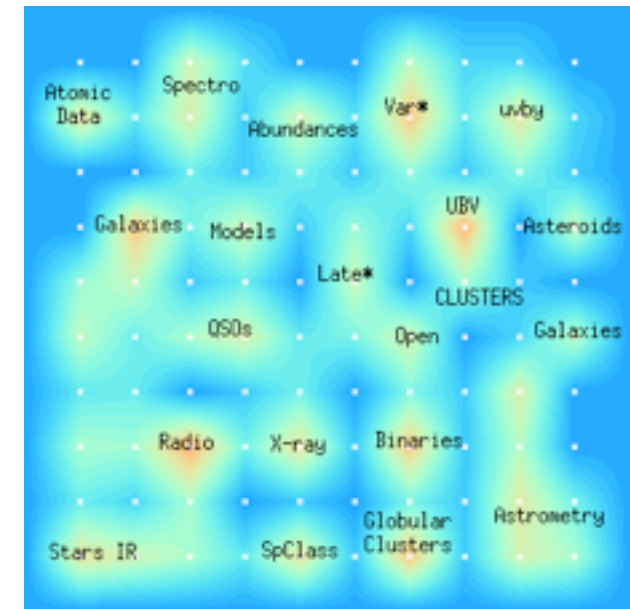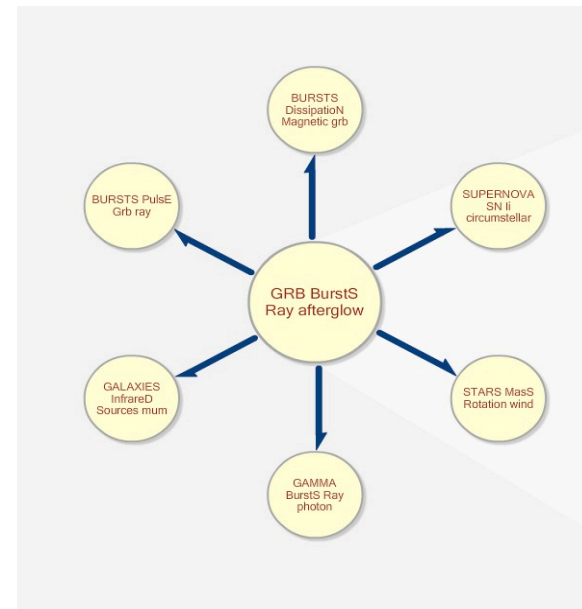
Semantic Search

- enable use of Knowledge Base in search expansion or narrowing

- use keyword extraction techniques to enable faceting and classification

- exploit usage data for ranking and recommendations

The ADS Linked Data Cloud

What can we do with this

- fulltext mining, log analysis allows us to create the needed links (keywords, citations, reads)

- clustering algorithms applied to citation and/or keyword co-occurrence networks detect coherent topics, classify documents

- co-readership and co-citation networks can be used to implement new query modes

- enhanced links to data allow the use of bibliographies to find relevant science products

Demo

- <u>Topic Search</u>

- <u>Data Search</u>

- <u>Keyword Topic Cluster</u>

- <u>User-based Facets</u>