Scholarly Communication Disruption and Transition

CS 431 – April 23, 2008

Carl Lagoze – Cornell University

Acknowledgements:

Les Carr
Herbert Van de Sompel
Tim Brody
Paul Ginsparg
Sandy Payette

Scholarly Communication vs. Popular Publishing

- Small, uniform author & reader community
- Authors and readers often the same
- Reliance on volunteerism and "community responsibility"
- Short-term readership interest

- Diverse and relatively large author & reader community
- Distinction between authors and readers
- Money and fame are motivating factors
- Interest often persists

Why do scholars publish?

- It is the tangible product of our work
- Our funders expect it – big publication lists always look good on reports
- It is our responsibility to our colleagues
- It is good for our egos
- It is the/a key to tenure, promotion, and hiring

(Very) short history of scholarly communication

- Pre-history: Scholarship through personal communication
- 1665: first scholarly journal
  - From face-to-face communication to more open accessible system
  - Anselm Strauss: *social worlds* built on texts
- Late 20th century: Monopolization
  - Distortion of journal model
  - "Serials crisis"
- 1990's: Digital Emergence
  - Web, E-journals, e-Print archives, institutional repositories
  - Reassertion of democratization
  - Access uber alles
- 21st century:
  - Blogs, wikis (social networks, web 2.0 meets scholarship)
  - Data centric

PHILOSOPHICAL TRANSACTIONS:

GIVING SOME

ACCOMPT

OF THE PRESENT

Undertakings, Studies, and Labours

OF THE

INGENIOUS

IN MANY

CONSIDERABLE PARTS

OF THE

WORLD.

Vol I.

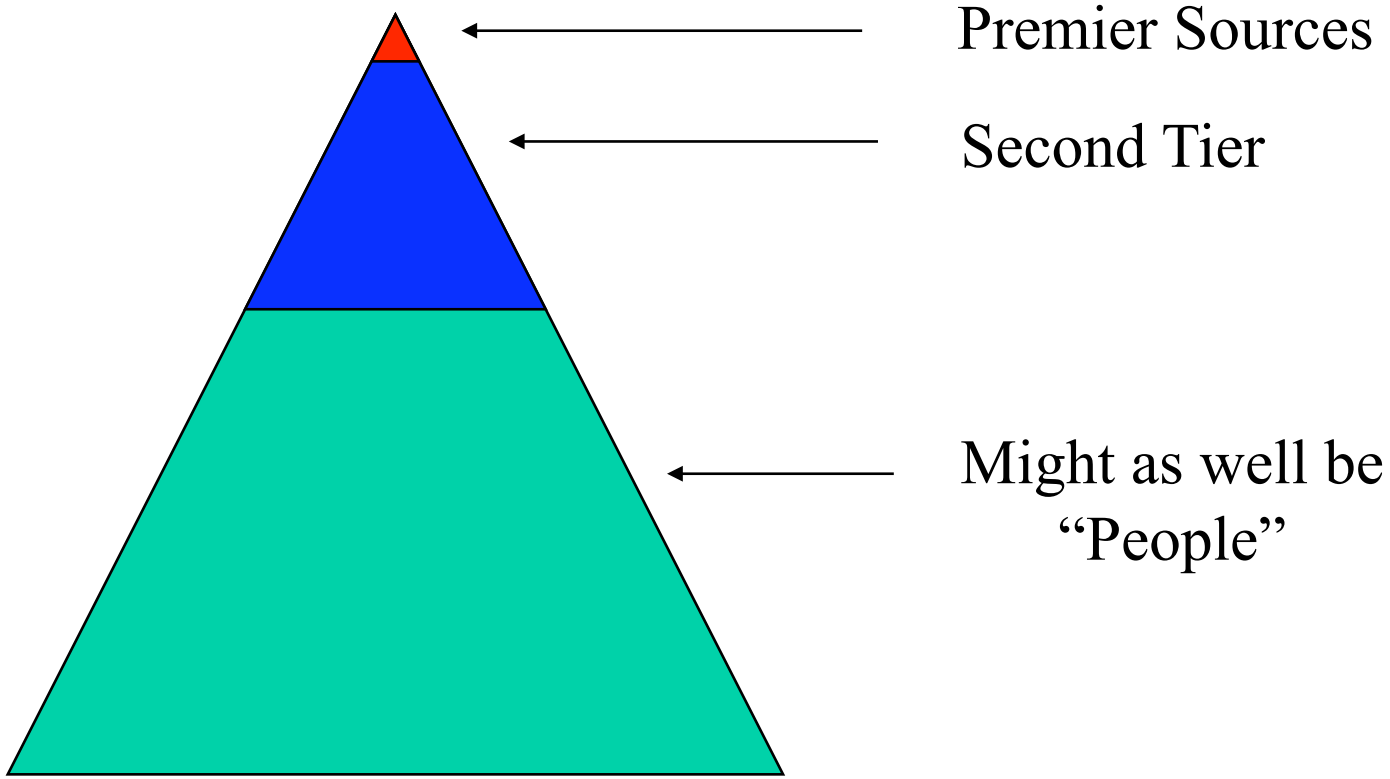
For *Anno* 1665, and 1666.

In the *SAVOY*,

Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*,

Printers to the *Royal Society*.

Scholarly publishing is extremely hierarchical



Peer Review

- Basis of quality in the system
- Almost entirely volunteer
- Blind or visible

Establishing Premier Journals – Citation Analysis

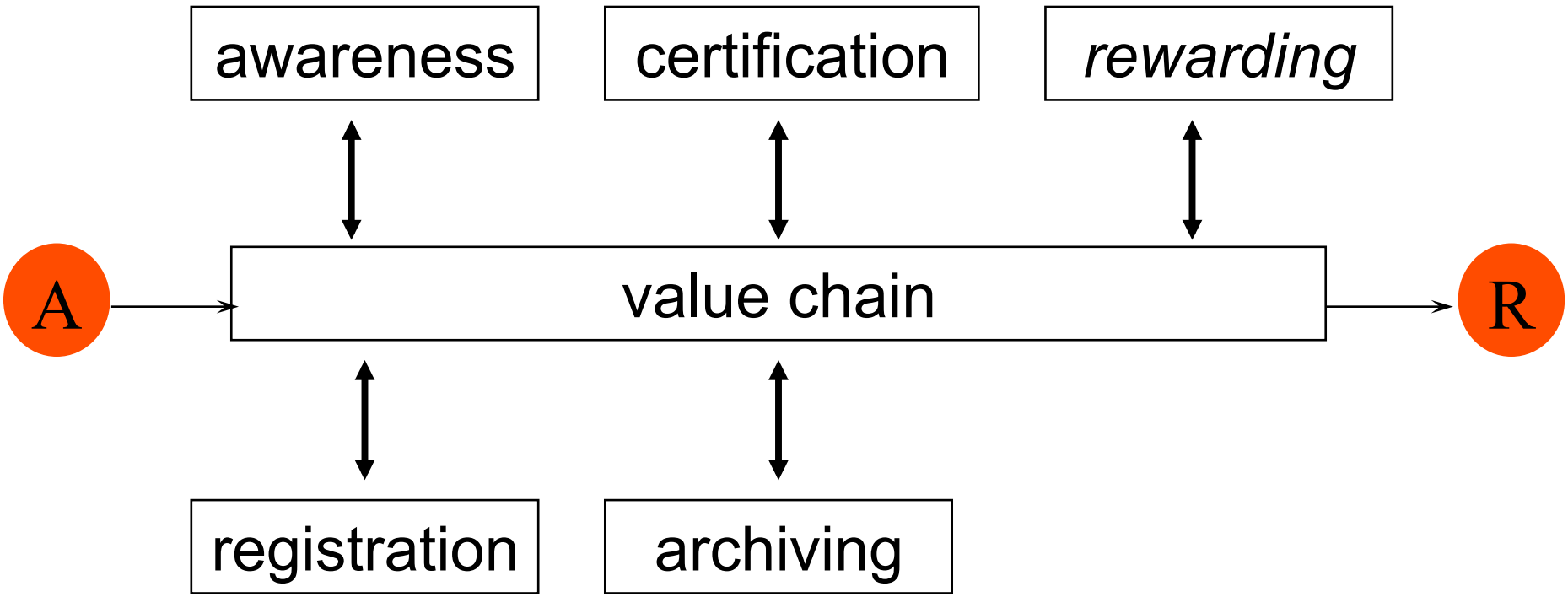
- A **citation** is a reference from one work to another [as a hyperlink: a citation link]
- Citation Graph – nodes are works, vertex is citation
- Citation analysis uses citation relationships to analyse patterns in research
- 'Bibliometrics'
  - (study of patterns in literature)
- Eugene Garfield
  - ISI Science Citation Index (SCI) identify "hottest" journals

Functions of scholarly communication

- **Registration** – to establish intellectual priority
- **Certification** – to certify quality and validity
- **Awareness** – to ensure accessibility
- **Archiving** – to endure availability for future use
- **Rewarding** – for tenure, promotion, compensation

(Roosendaal & Geurts)

Value chain perspective of scholarly communication system

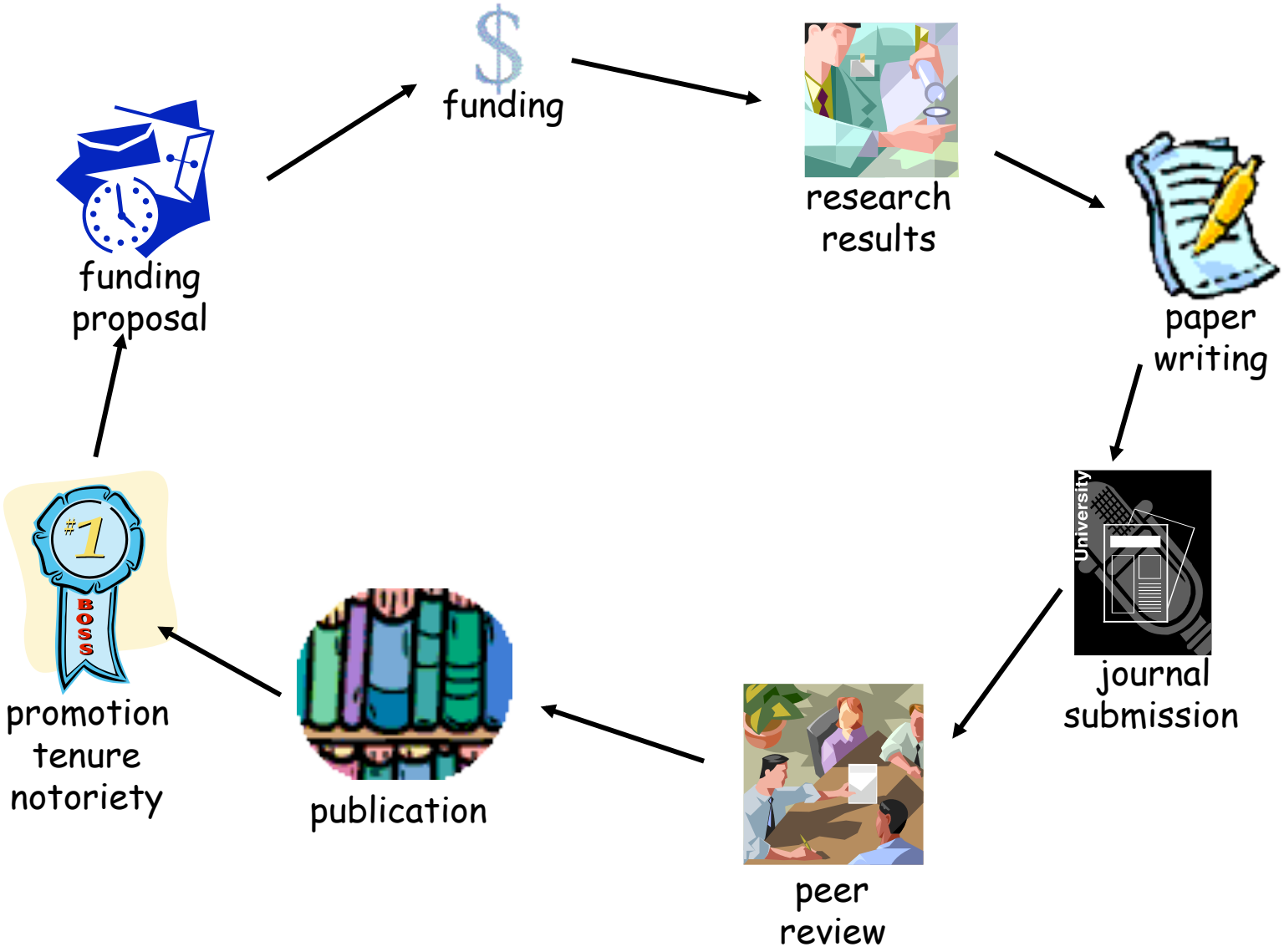


Traditional journal system integrates functions

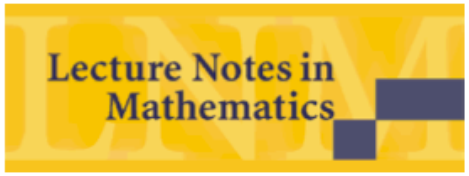
- Provides certification (usually via peer review)
- Accepted status of journals provides for rewarding
- Libraries provide archiving (and shoulder additional cost)

- And, in fact, locks out anything that doesn't pass through this path

How the system works



Consent to Publish



Title of the Multi-Author Book/Volume:

Volume Editor(s) name(s):

Title of Contribution:

Author(s) full name(s):

Corresponding Author's name, address, affiliation and e-mail:

§ 1 Rights Granted

The copyright to the Contribution identified above is transferred to Springer-Verlag GmbH Berlin Heidelberg (hereinafter called Springer-Verlag). The copyright transfer covers the sole right to print, publish, distribute and sell throughout the world the said Contribution and parts thereof, including all revisions or versions and future editions thereof and in any medium, such as in its electronic form (offline, online), as well as to translate, print, publish, distribute and sell the Contribution in any foreign languages and throughout the world.

Springer-Verlag will take, either in its own name or in that of the Author, any necessary steps to protect these rights against infringement by third parties. It will have the copyright notice inserted into all editions of the Work according to the provisions of the Universal

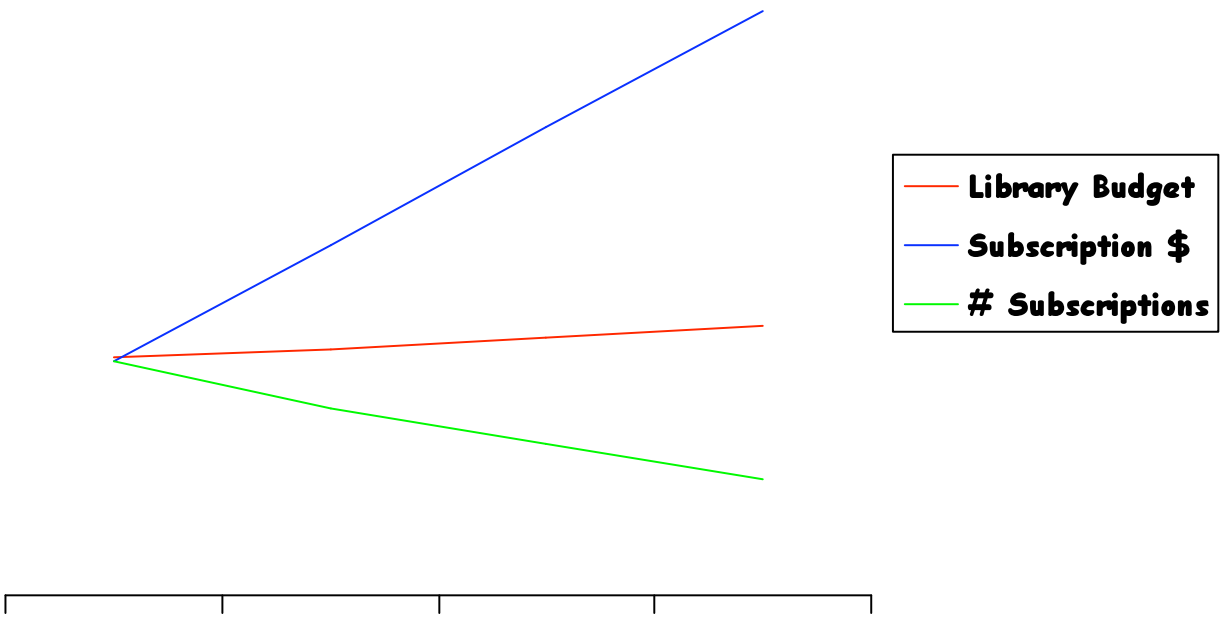
Who are the role players

- Scholars
  - Faculty
  - Researchers – Commercial, Academic, Government Labs
- Publishers
  - "Big" for-profits: Elsevier, Springer-Verlag
- Learned and Professional Societies
  - ACM, APS, AMS
    - Publishing operations often subsidize other operations
    - Some are hard to differentiate from for-profit publishers – e.g., IEEE
- Libraries
  - In paper system they were the sole distribution point for publications

Issues and Changes

- Exponentially increasing amount of information produced by scholars
- Growth in both dimensions
  - Horizontal
    - Increased specialization
    - New and more specialized journals
      - 5000 peer reviewed journals in education research
  - Vertical
    - Diminish single source reliance
    - Facilitate multi-uses for single source
- Compressed time for "relevance" of results, increased demand for rapid delivery
- Changes in the type of publication
  - demand for data availability

Broken Economics



Some reflections on subscription prices

- Average journal subscription price has gone up 7-10%/year over the past 10 years
  - 1986-2002 US CPI increased 57%, research library journal subscription budget increased 227%
  - Average journal has increased 186% (outstripped inflation 4 times)
- Some journals have gone up 20-40% of the past 5 years!!!
- Some journals cost 5K-10K per year
- Many societies have raised subscription prices 20-25% over the past several years
  - "Catch up" to the private publishers
  - Fund research into digital initiatives
  - Cover the rest of their operations
- Elsevier's price rise per year equates to one less faculty member per year (according to Bill Arms)
- http://oap.comm.nsdl.org/10most.html

Assumptions in current scholarly publishing system

- Publications are difficult to produce
- Publications are difficult to distribute
- Readership is by closed community
- Quality assessment is by closed community
- Archiving and management is by closed community

Some "side effects" of the current system

- **Rich get Richer!**
  - Best known scholars have an advantage in peer review system
  - Riches institutions in richest countries can best afford journal prices
  - High prestige journals are self-sustaining due to SCI factors
- Global scholarly divide worsens
  - Research institutions in developing countries can't afford subscriptions
  - Intellectual capital flees
- Hierarchy gets more stratified
  - Unpublished papers disappear
  - Entry into the system is difficult

Where are the costs in the print system

- Publishers
  - Copy-editing
  - Production
  - Administration of review system
  - Production
  - Distribution
- Libraries
  - Cataloging
  - Preservation
  - Binding
  - Shelving

Economics have changed!

- Distribution in electronic system is basically free
  - Fundamental assumption of paper system is eliminated
  - "Publishing" by everyone should be encouraged and supported
- Services need to be disambiguated from distribution
  - Free distribution doesn't mean that there isn't an economic model
  - Systems like review, filtering, awareness can be built on top of a free distribution system

Signs of Change - Publishers

- Electronic versions of existing journals
- Licensing arrangements to libraries
- Problems
  - License bundling
    - Inflate costs and maintain economic model
    - Force libraries to subscribe regardless of interest
  - Longevity dependent on license continuity
- Specialty portals
  - Scirus (http://www.scirus.com)

Signs of Change – Libraries & Professional Societies

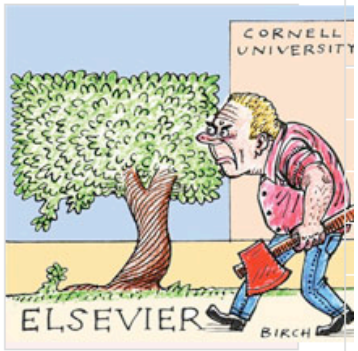
- Realities
  - Many professional societies and journals are "Mom & Pop" operations
  - Technical and economic cost of electronic publishing is often prohibitively high
- HighWire Press – http://highwire.stanford.edu
- Solution
  - Highwire acts as a brokering service to provide electronic publishing technology for small professional societies and journals
  - Pooling technology allows creation of higher level services (e.g., reference linking amongst journals)

News

Nature **426**, 217 (20 November 2003) | doi:10.1038/426217a

Cornell axes Elsevier journals as prices rise

Jonathan Knight, San Francisco



A top US research university is set to cancel its subscriptions to several hundred scientific journals published by Elsevier in January, in response to spiralling subscription costs.

The decision by Cornell University in Ithaca, New York, released in a statement last week, caught the attention of library officials at other US research universities who say that they may follow suit.

Netherlands-based Elsevier, which owns a quarter of the global market in scientific and technical journals, played down the importance of the move. Eric Merkel-Sobotta, its head of public relations, declined to comment on the company's negotiations with Cornell, but said that the firm is working hard to accommodate all of its customers. He added that Elsevier's subscription rates are rising by less than 7% annually, an increase that is necessary to cover the cost of expanding journal content.

Like other large publishers, Elsevier offers 'bundled' subscriptions to its journals along with electronic access. Many institutions have signed such agreements to gain access to large numbers of journals.

"This started out as a very good deal for universities," says Ted Bergstrom, an economist at the University of California, Santa Barbara, who studies journal pricing. But the cost of such arrangements has risen faster than the rate of inflation, and economic woes have put library budgets under pressure.

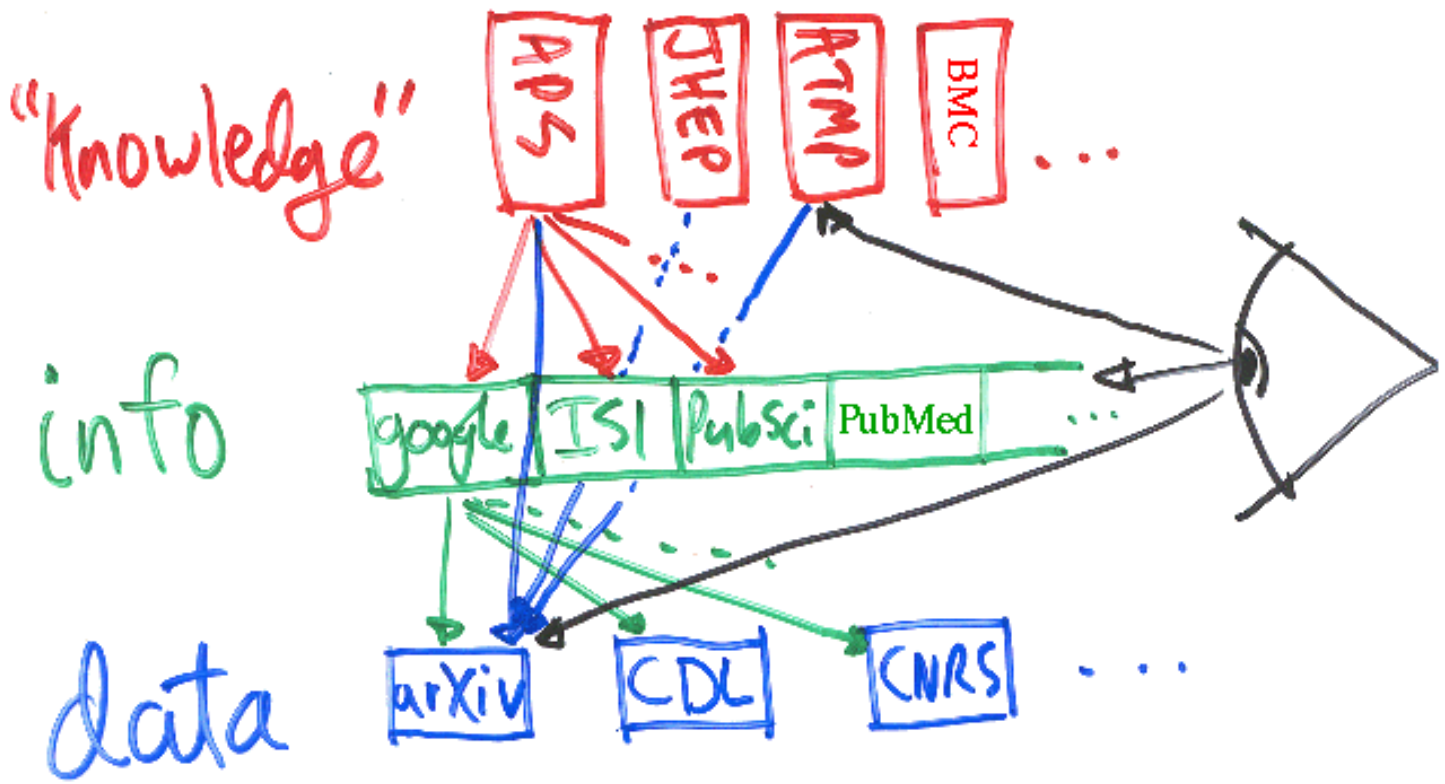
At Cornell, the increases have forced the library to cut back on its non-bundled titles — but bundles cannot be touched without abrogating the original pricing deal.

Cornell's deal with Elsevier, now priced at $1.7 million, consumes a fifth of the university's total periodical budget. When the library tried to cancel individual Elsevier titles, university officials say, the prices of the remaining titles increased significantly, offsetting any savings. "To save a little, you have to cancel a lot," says Cornell's associate collections librarian, Ross Atkinson. Cornell will now return to a title-by-title plan with a vastly reduced number of journals, he says.

Cancellations by other universities are also likely, says Duane Webster, director of the Association of Research Libraries in Washington DC. "Cornell is just the

Signs of Change - Publishers

- Electronic Journals
  - D-Lib Magazine – http://www.dlib.org
  - Ariadne - http://www.ariadne.ac.uk/
  - First Monday - http://www.firstmonday.org/
  - Journal of Electronic Publishing - http://www.journalofelectronicpublishing.org/
- The economic models are not established



A Future with a View

Acks. P. Ginsparg

What are the implications of this model?

- A marketplace of ideas
- People choose appropriate entry points into the system
  - Troll for free at the lowest layers
  - Pay for guided entry at upper layers
- Exposure of the "long tail"
- Money can be made for synthesizing information
- Standards for interchange amongst layers are important (e.g., OAI-PMH)

Signs of Change – Open access movement

- Free, immediate, permanent, full-text, online access, for any user, web-wide, to digital scientific and scholarly material, primarily research articles published in peer-reviewed journals.
- Various declarations
  - Budapest - http://www.soros.org/openaccess/
  - Berlin - http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html
- Author "self-archiving" – storage of pre-print or post-print in own or "institutional repository"
- Significant effect on publishers
  - Over 90% now allow some form of author self-archiving

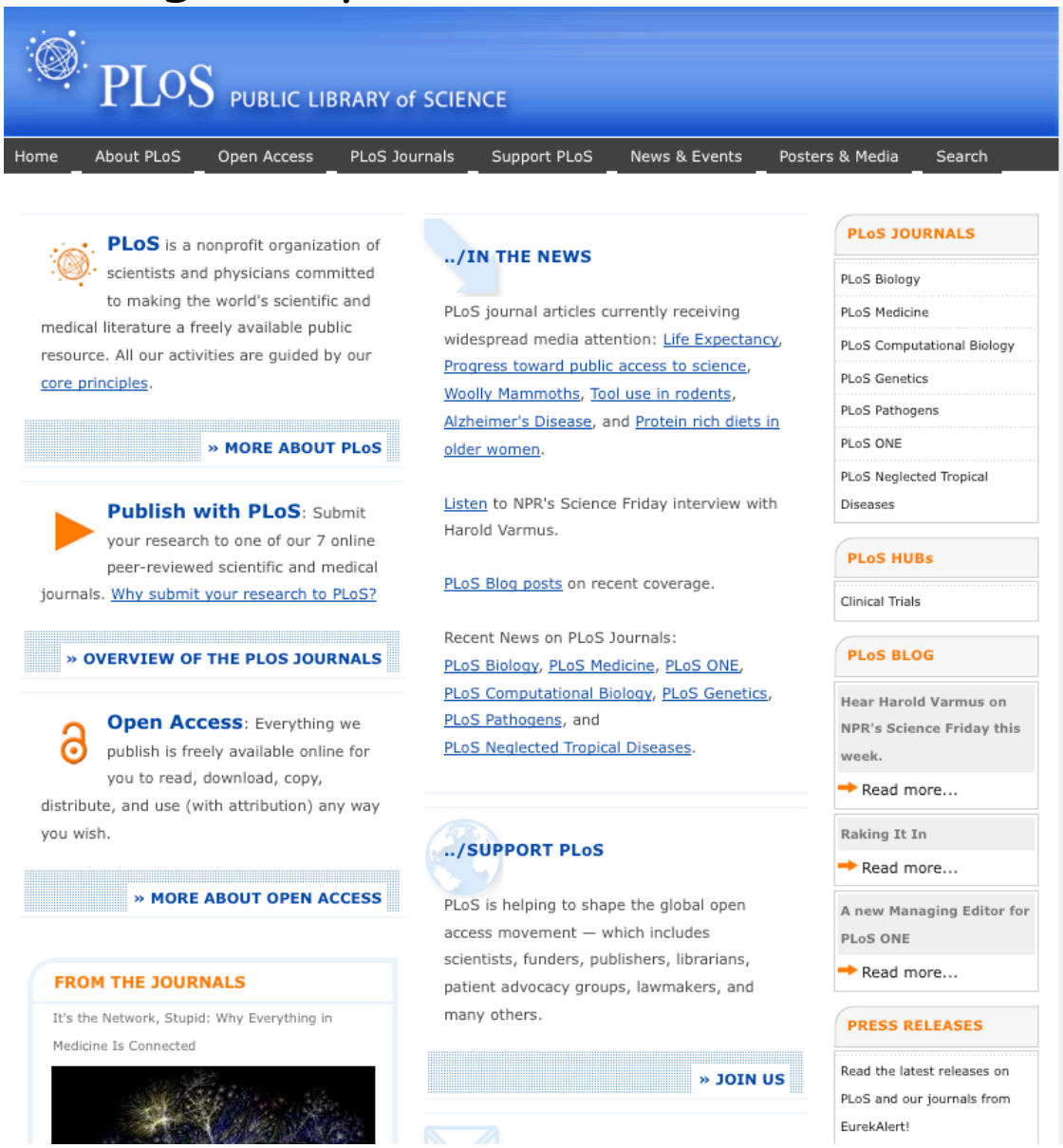
Signs of Change – Open Access

- Eprint respositories
  - Author-self archiving gives scholars control over their intellectual output
  - Harnad's "subversive proposal"
  - Direct descendant of traditional pre-print sharing in print form among scholars
- Examples
  - arXiv – http://arxiv.org
  - ePrints – http://www.eprints.org
  - California Digital Library scholarly publishing archive - http://repositories.cdlib.org/

Signs of Change – Institutional Repositories

- Institution-based
- Scholarly material in digital formats
- Cumulative and perpetual
- Open and interoperable
- DSpace (http://www.dspace.org)
  - Institutional Repository for MIT faculty's digital research materials
  - MIT Libraries - Hewlett Packard Research Labs collaborative development project
  - Open Source system
  - Federated system
  - Preservation archive

Signs of Change – Open Access Journals



Signs of Change – The Web and Computer/ Information Science

- Automatic creation of traditional journal services
- CiteSeer – http://citeseer.ist.psu.edu/
  - Selective web crawling to gather CS resources
  - Heuristics and AI techniques to establish services
    - Searching
    - Reference linking
- Follow-on by commercial search engines
  - Google scholar – http://scholar.google.com
  - Windows Live Academic - http://academic.live.com/

Digitometric/Infometric Analysis

- Bibliometrics for the online age
- Couple citation analysis with Web analysis
  - (how many times has x been accessed?)
- Similar to readership studies, but easier to survey and more comprehensive
  - (though subject to the same problems of copies being re-distributed, multiple accesses etc.)

Predicting Citation Impact

- The Web gives us access to new metrics
  - Download/access frequency
- Can early-day 'download' frequency give an indication of longer-term citation frequency?
- Not all citations are equal
  - Understanding the nature of citations
  - Structural and contextual analysis

Considering Peer Review

- In general agreement that peer review should be an important quality factor
- Alternatives
  - Combine with other factors
    - Weighted PageRank
    - Various other graph analysis methods
  - Recognize other quality ratings
    - Usability by other communities and contexts
    - Factors for data "quality" are different

Setting More Ambitious Goals

- But, we've only created an electronic equivalent of the paper-based system.

PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

investigated by Solid Phase Micro Extraction-Gas Chromatography/Mass Spectrometry (SPME-GC/MS). Small amounts of these aerogels and of the adhesive tape (0.5 - 1.7 grams) were introduced into 40 mL glass vials and heated at ~ 85° C for twenty minutes. Then a SPME fibre (divinylbenzene-carboxen-polydimethylsiloxane) by Supelco [Sup] was exposed for five minutes into the head space in order to sample the outgassing substances.

One sample of SP-50 ('04 configuration) has also been prepared in slightly different conditions. The aerogel tile has been cut into smaller fragments with respect to the previous experiments in order to sample a bigger amount (~2.4 grams) and the SPME fibre was exposed for 20 min. (Fig. 1b). Procedure blanks have been carried out as well.

After exposure, the SPME fibre was introduced into the injection port, maintained at ~ 250° C, of a gas chromatograph (Trace GC ultra, Finnigan-Thermo) coupled with an ion trap mass spectrometer (Polaris Q, Finnigan-Thermo). The MS ion source temperature was kept at 250°C. The mass spectrometer was operating in the EI positive mode (70 eV) with a mass range of 45-650 *m/z* (mass over electric charge). For the gas chromatographic separation a 5% diphenyl-95% dimethyl polysiloxane column (SPB-5, Supelco) (30m × 0.25mm i.d., 0.25 µm film thickness) was used in splitless mode. The chromatographic conditions were: 40°C (5 min.), 5°C/min., 250°C (20 min.). The carrier gas (He) was used in constant flow mode at 1.0 ml/min.

2.2 Results

In the chromatograms of all aerogel samples (Figs. 1-2 and Table 2) a low molecular weight siloxane has been revealed at RT 2.83. This compound is probably related to the chemical treatment of the aerogel used to confer hydrophobic properties.

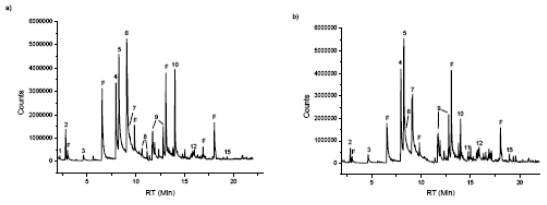
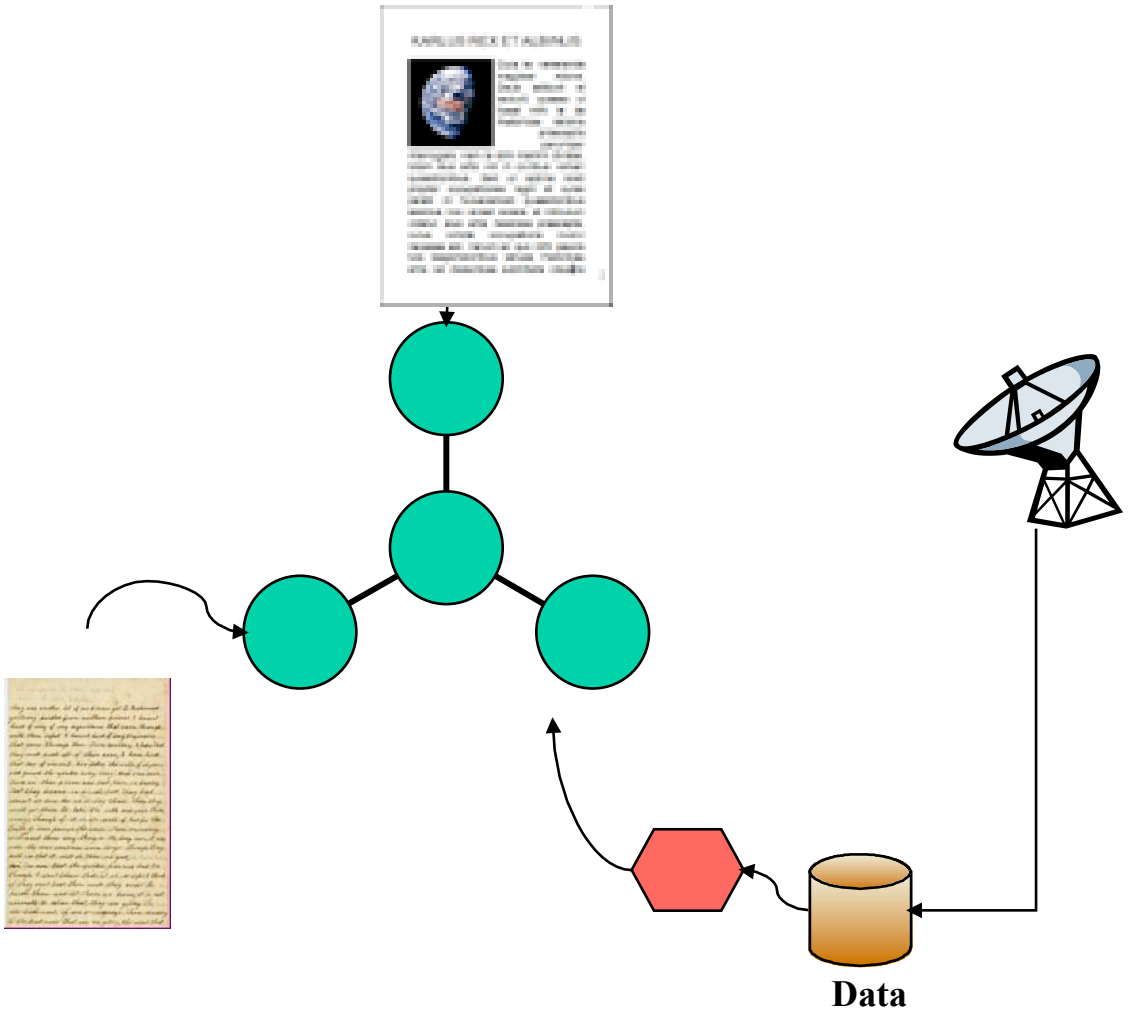


Fig. 1: Chromatograms of Silicon aerogel samples. The two samples of SP50 a) unused ('05 configuration) and b) used ('04 configuration) have been analyzed with the same SPME-GC/MS technique but two slightly different procedures, read text for more details. The main identified compounds are listed in Table 2. Peaks of siloxanes, due to the SPME fibre, are labelled F in the chromatograms.

The new "**information unit**"

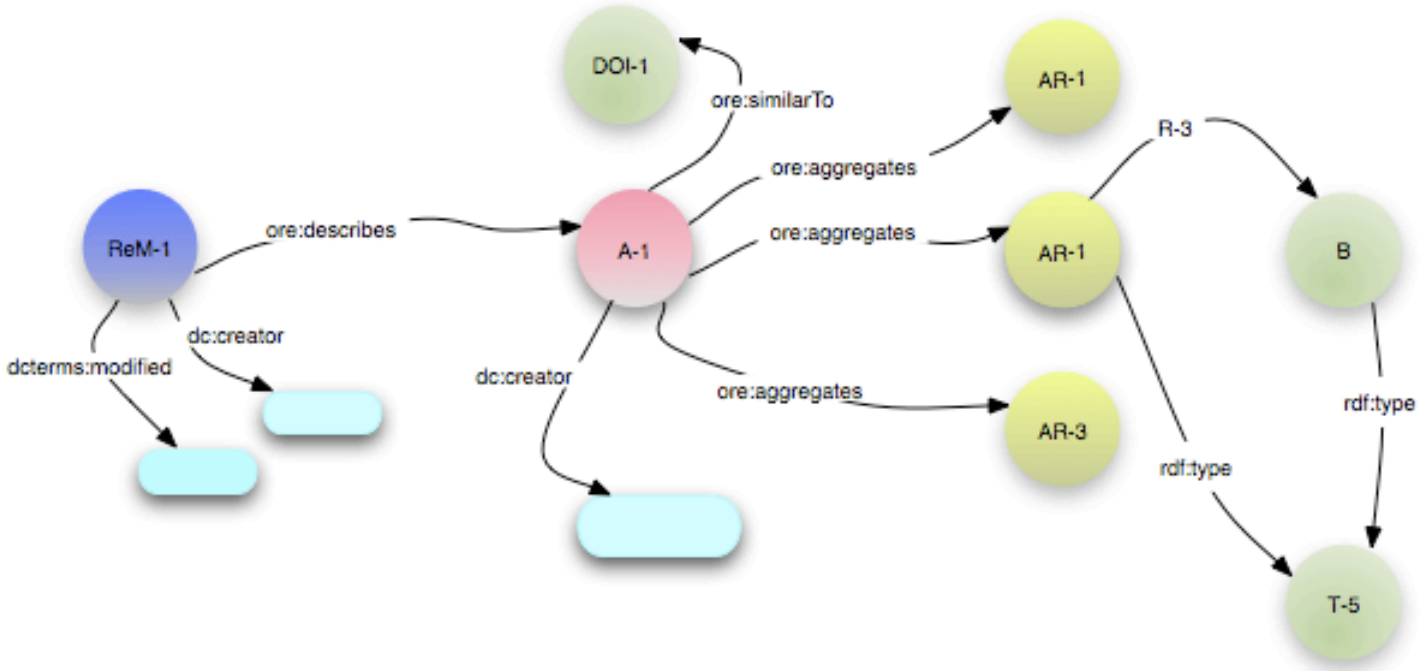
- Documents
- Text
- Data
- Simulations
- Images
- Video
- Computations
- Automated Analyses

Aggregations



Aggregation Model

- Identification, description, deconstruction, re-use

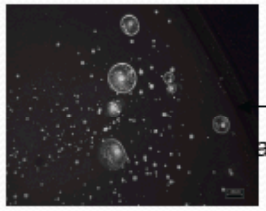


Ideal - Scientific Publication Packages

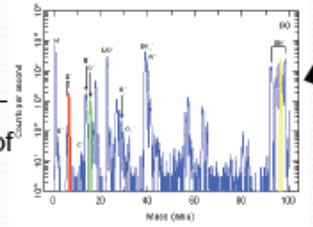
	Area	Mean	S.D.	X	Y	Mode	Length	Major	Minor	Angle	Int.Den	Back.	Min	Max
1	0.01	208.2	88.14	0.34	0.06	253	0.34	0.11	0.08	102.7	0	0	35	253
2	0.01	206.8	89.14	0.17	0.07	253	0.34	0.1	0.08	17.57	0	0	35	253
3	0.01	212.9	84.54	0.26	0.11	253	0.37	0.11	0.1	158	0	0	35	253
4	0	190.4	98.85	0.07	0.1	253	0.21	0.07	0.05	76.53	0	0	35	253
5	0.03	228.8	68.54	0.67	0.38	253	0.75	0.24	0.15	154.8	0	0	35	253
6	0.09	240.7	50.36	0.34	0.48	253	1.24	0.38	0.3	95.89	0	0	35	253
7	0.08	240.1	51.46	0.59	0.59	253	1.18	0.35	0.28	81.38	0	0	35	253

External Database

Derived_from



analysis_of

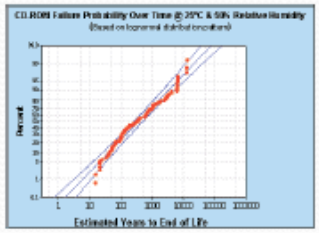


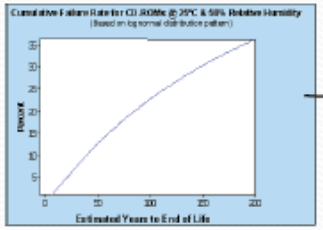
graph_of

Average LE = 1/T exp –(A –B/T)

derived_from

refers_to





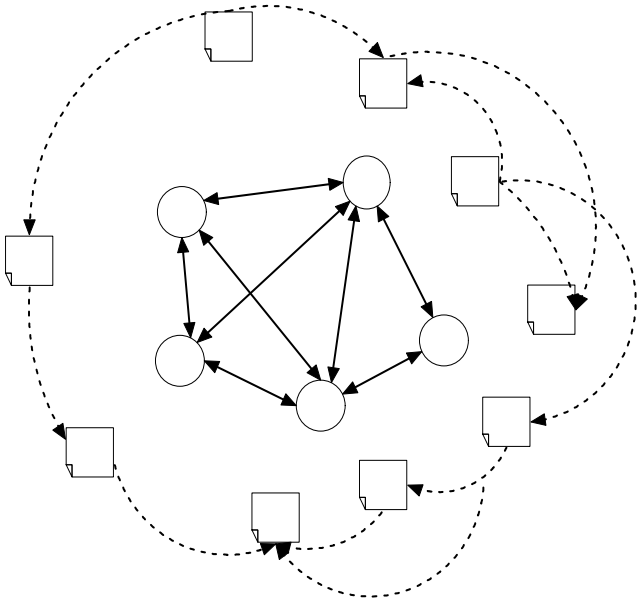
refers_to

Slattery, O., Lu, R., Zheng, J., Byers, F., Tang, X. "Stability Comparison of Recordable Optical Discs- A study of error rates in harsh conditions," Journal of Research of the NIST, 109, 517-524, 2004

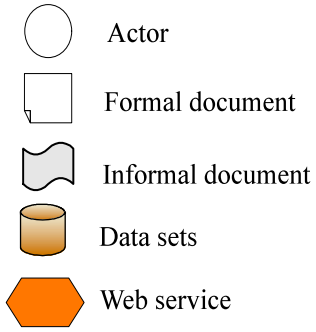
RDF Package

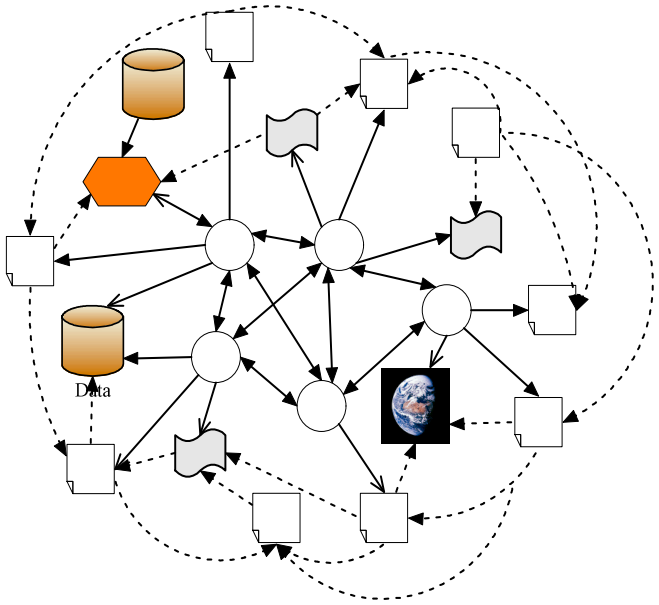
Title
Creator
Description
Type
Discipline
Date.Published
License

Integrating Social networks and bibliographic networks



Disconnected networks:
- **formal publication network**
- **social network (actors)**





Hybrid network
- **documents (formal and informal)**
- **data**
- **services**
- **actors**