

Scholarly Communication Disruption and Transition

CS 431 - April 26, 2006

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

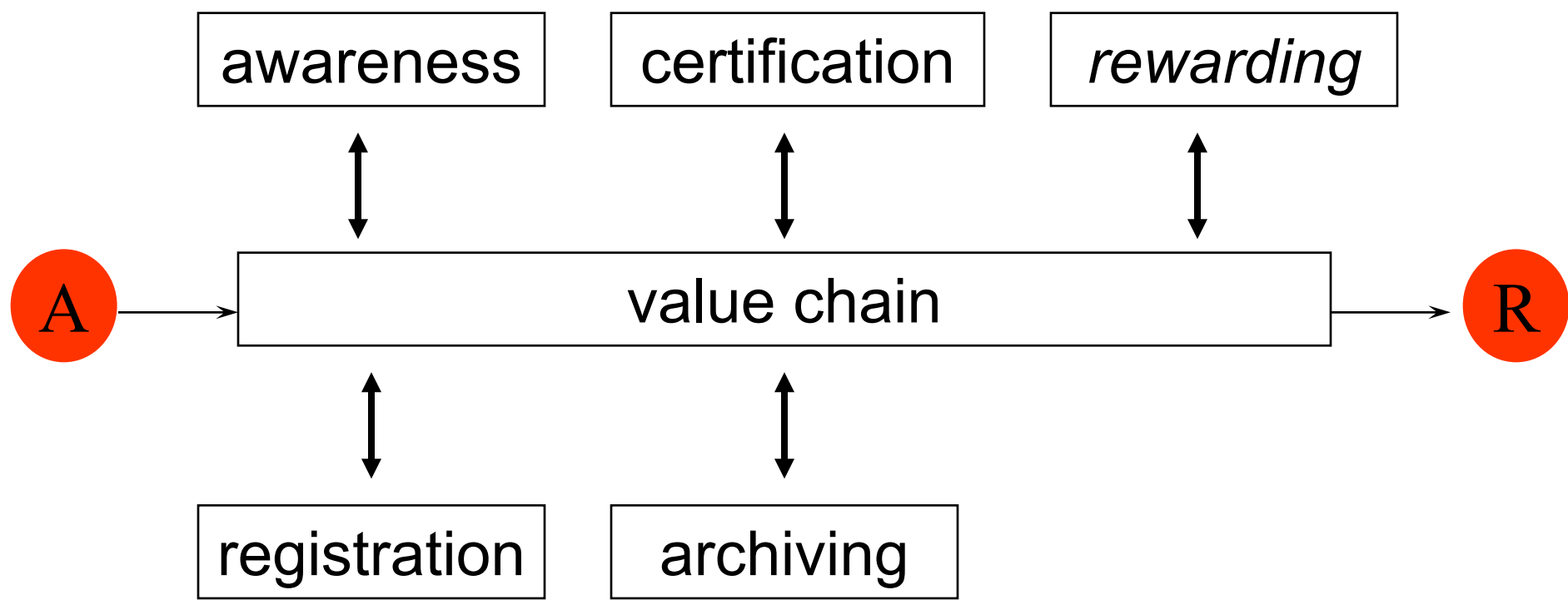
- Pre-history: Scholarship through personal communication
- 1665: first scholarly journal
 - From face-to-face communication to more open accessible system
 - Anselm Strauss: *social worlds* built on texts
- Late 20th century: Monopolization
 - Distortion of journal model
 - “Serials crisis”
- 1990’s: Digital Emergence
 - Web, E-journals, e-Print archives, institutional repositories
 - Reassertion of democratization
 - Access uber alles
- 21st century: ??

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)

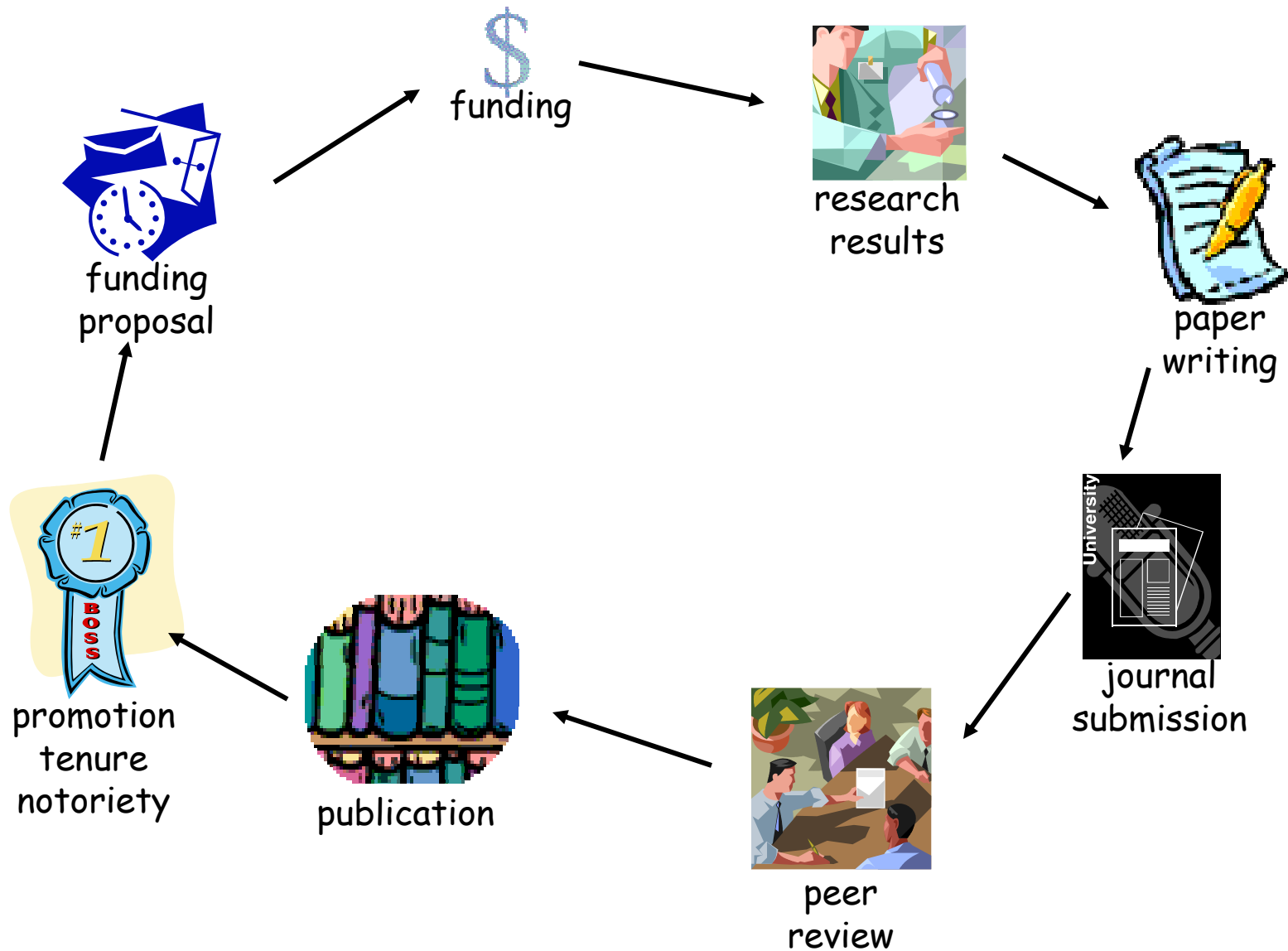
Value chain perspective of scholarly communication system



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



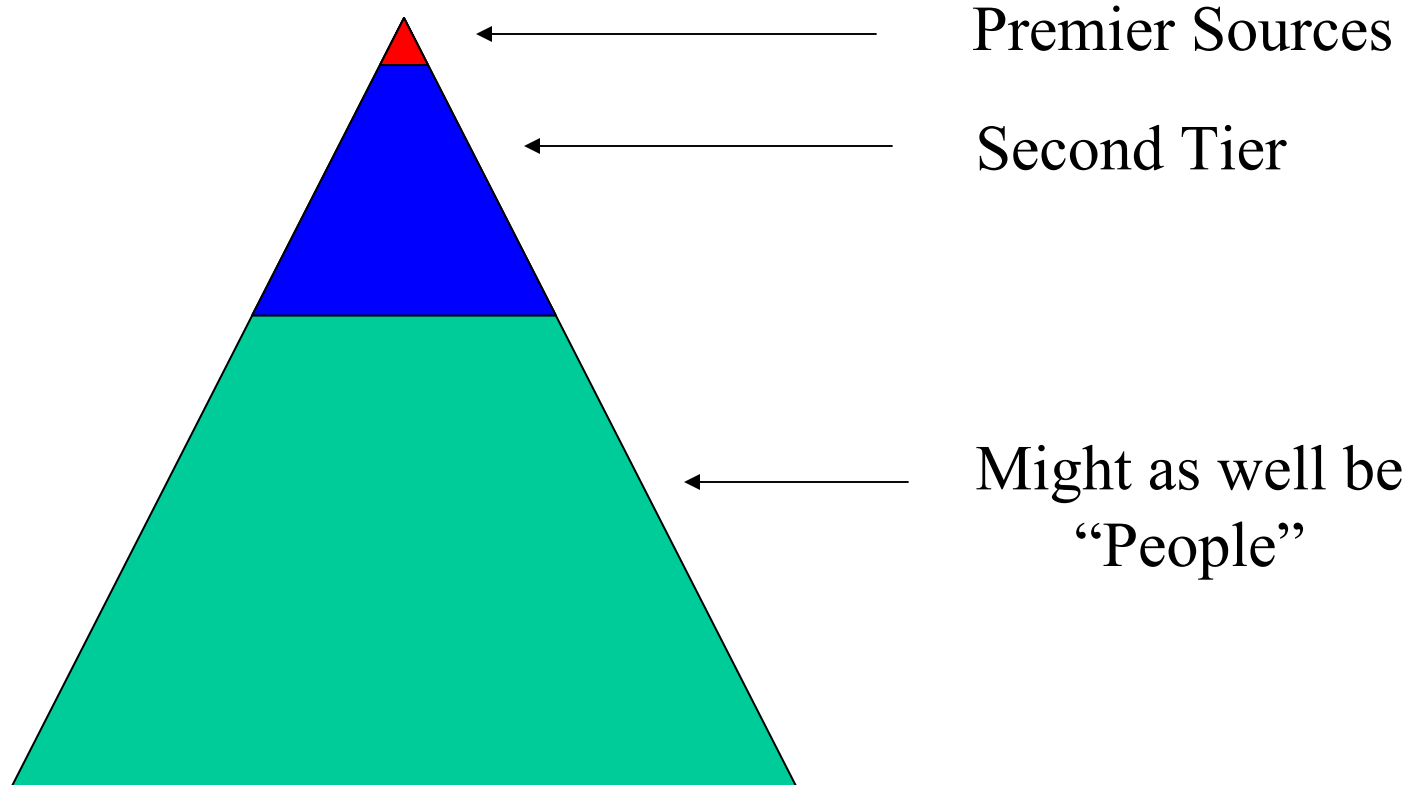
Peer Review

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

Who are the role players

- Scholars
 - Faculty
 - Researchers - Commercial, Academic, Government Labs
- Publishers
 - "Big" for-profits: Elsevier, Springer-Verlag (Kluwer)
- 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 the sole distribution point for publications

Scholarly publishing is extremely hierarchical



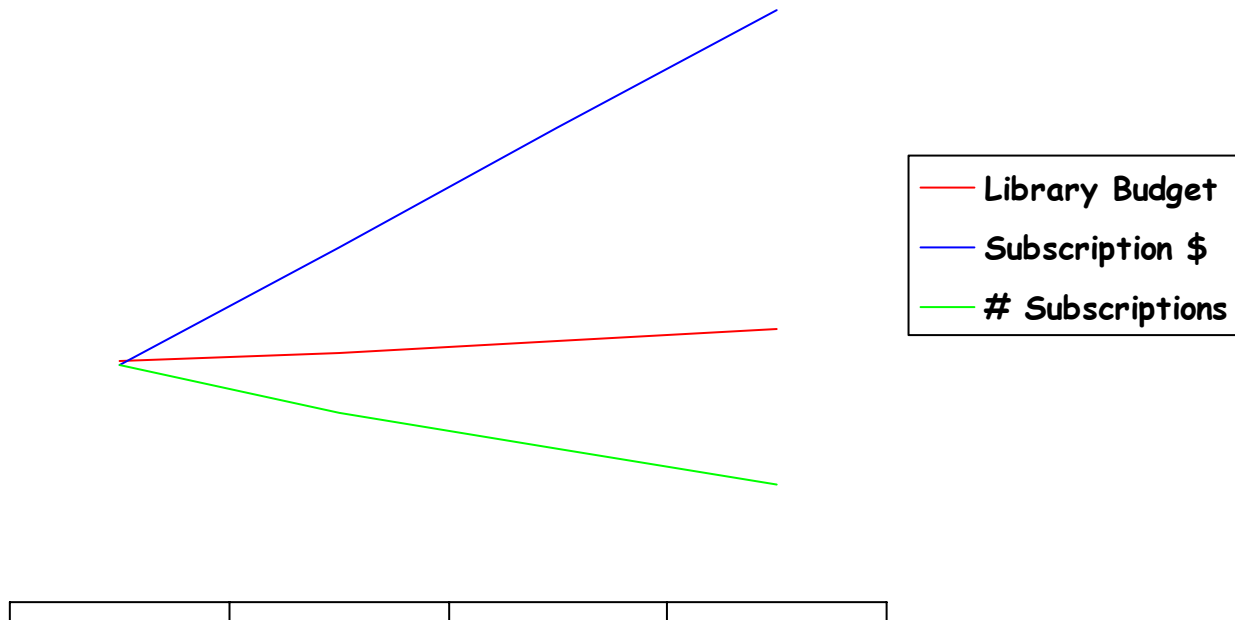
Establishing Premier Journals - Citation Analysis

- *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

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%
- 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.nsd.org/10most.html>

Assumptions in current scholarly publishing system

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

Some “side effects” of the current system

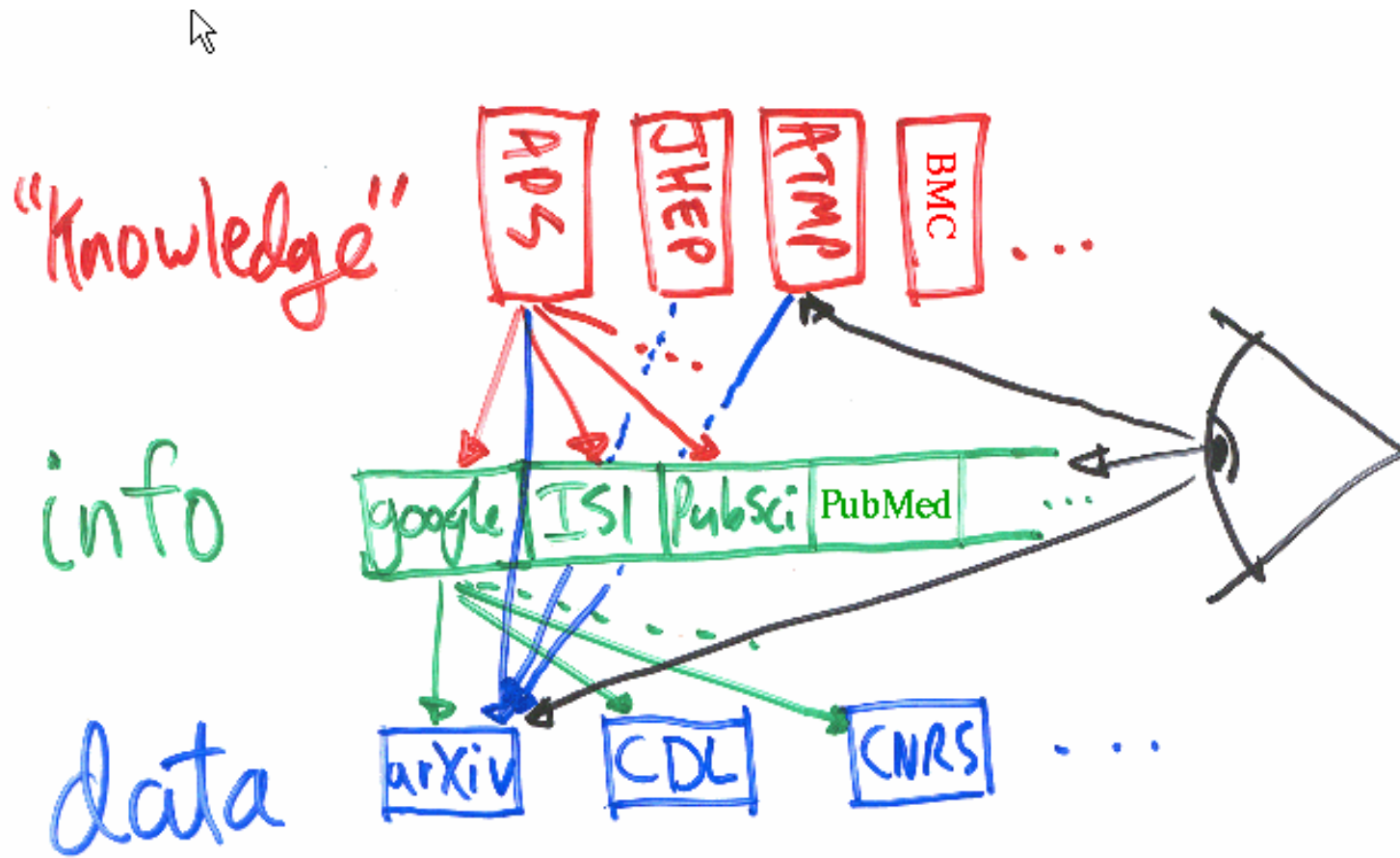
- *Rich get Richer!*
 - Best known scholars have an advantage in peer review system
 - Riches institutions in richest countries can best afford journal prices
 - High prestige journals are self-sustaining due to SCI factors
- Global scholarly divide worsens
 - Research institutions in developing countries 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



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 - Readers

... there's a sense in which the journal articles prior to the inception of the electronic abstracting and indexing database may as well not exist, because they are so difficult to find. Now that we are starting to see ... full-text showing up online, I think we are very shortly going to cross a sort of critical mass boundary where those publications that are not instantly available in full-text will become kind of second-rate in a sense, not because their quality is low, but just because people will prefer the accessibility of things they can get right away.

Clifford Lynch 1997

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 - Publishers

- Electronic Journals
 - D-Lib Magazine - <http://www.dlib.org>
 - Ariadne - <http://www.ariadne.ac.uk/>
 - First Monday - <http://www.firstmonday.org/>
- The economic models are not established

Signs of Change - Libraries & Professional Societies

- HighWire Press - <http://highwire.stanford.edu>
- Realities
 - Many professional societies and journals are "Mom & Pop" operations
 - Technical and economic cost of electronic publishing is often prohibitively high
- 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)

Signs of Change - Scholars

- Eprint repositories
 - 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/>
- Related Issues
 - Publisher agreements - some journals refuse to publish anything that has been posted as an eprint

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 - 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

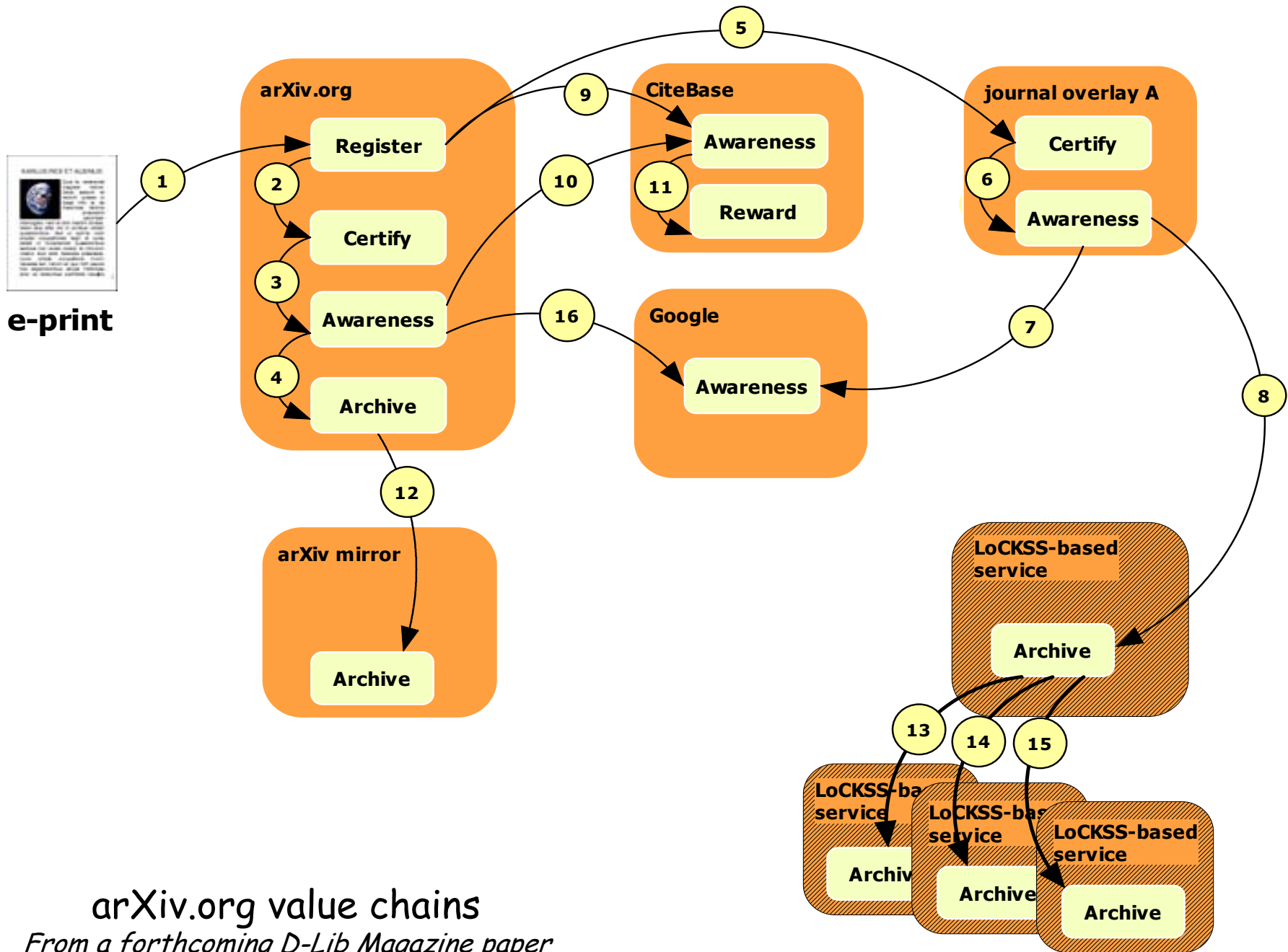
- But, we've only created an electronic equivalent of the paper-based system.
- The networked environment provides opportunities for more radical changes.
- While 'open access' is important, it should not be our only focus.
- Exploit new opportunities
 - Deconstruct the value-chain
 - Build communities around documents
 - Define new information objects

Institutional Repositories: A foundation of the new model



Repositories

- Typically textual materials
- Different levels of **Certification**: no review, internal review, curatorial decisions, peer-review, ...
- New value chains will emerge with these materials as their starting point

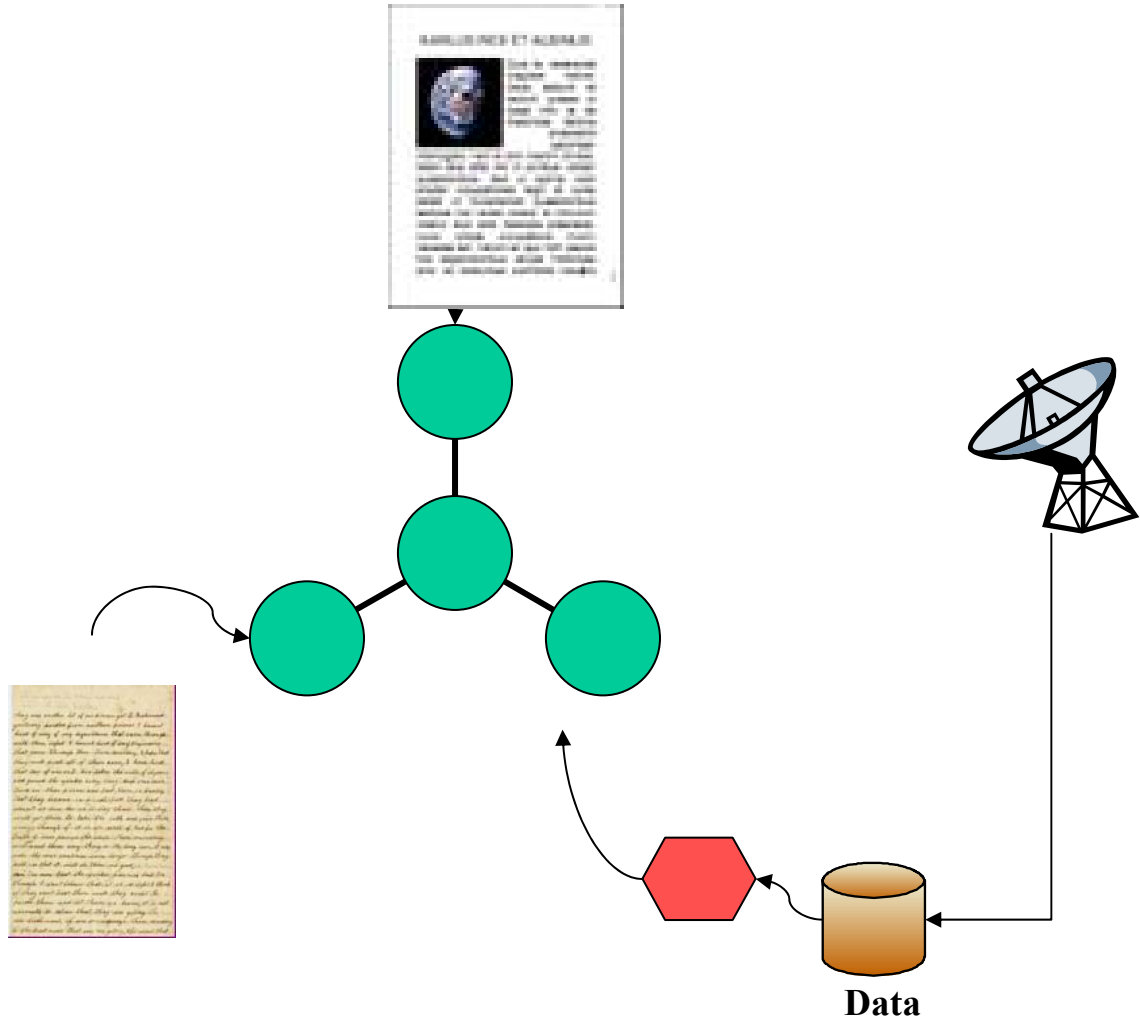


arXiv.org value chains
From a forthcoming D-Lib Magazine paper

The new "information unit"

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

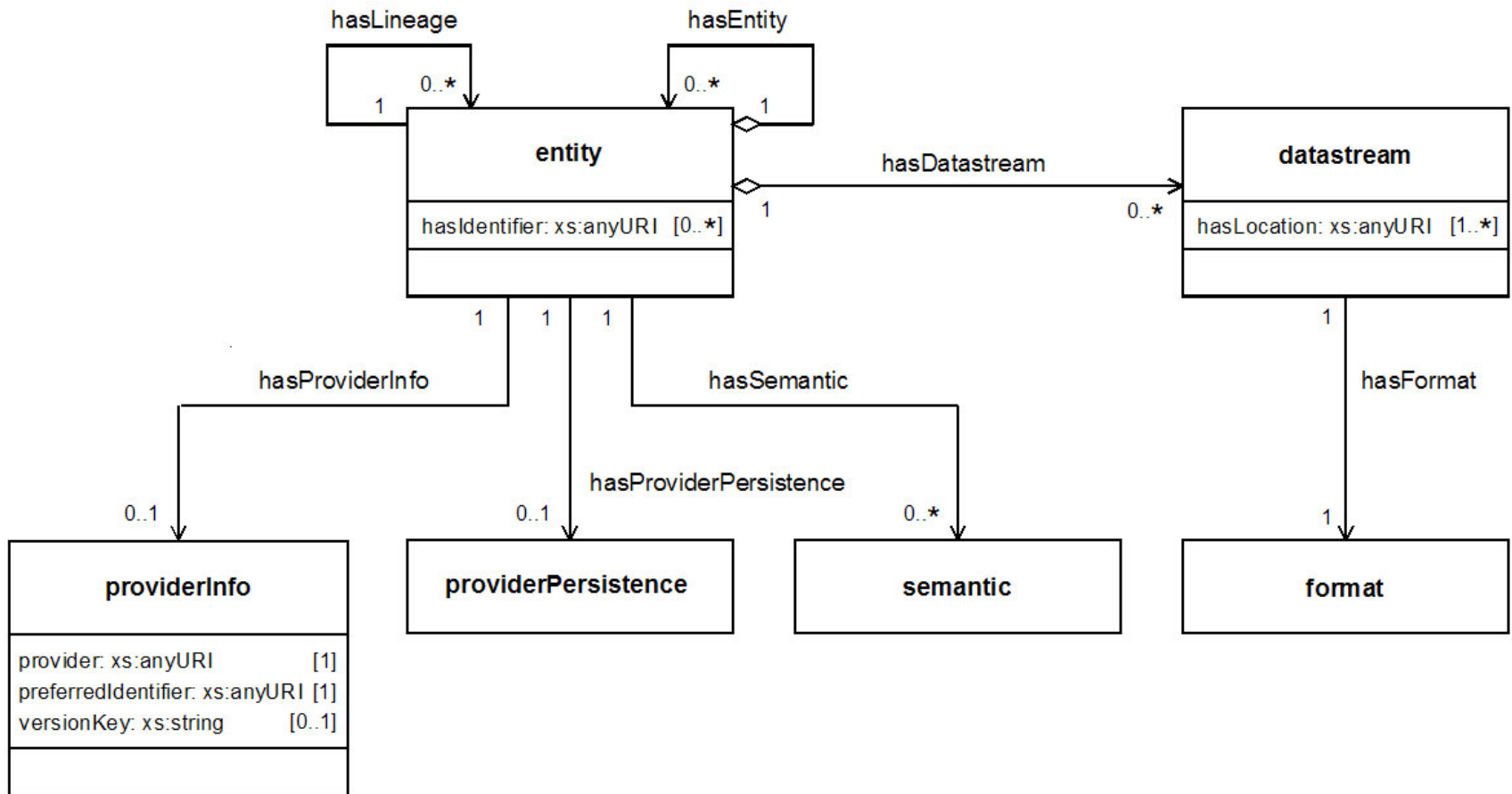
Aggregations



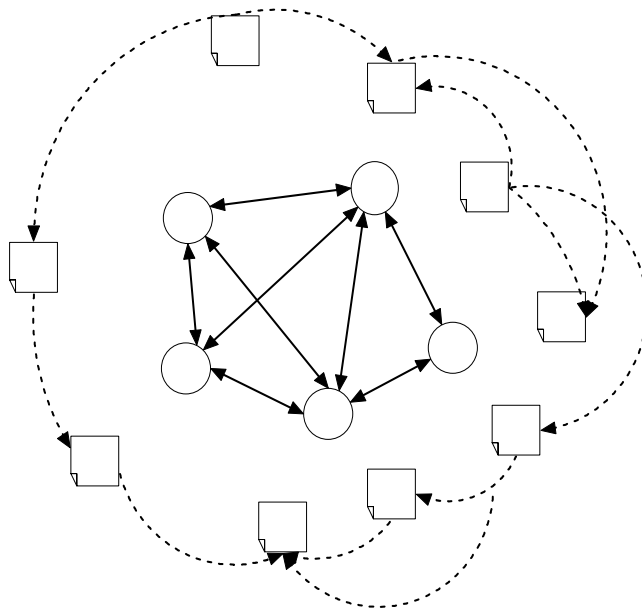
Model Core Requirement

- **Identity** independent of specific schemes
- **Lineage** relationships among objects
 - evidence of workflow for evidential citation
- **Semantics** associated with entities
 - facilitate service mapping
- **Recursion** for n-levels of entity containment
- Link to **concrete representation**
- Assertion of **persistence** levels

Data Model

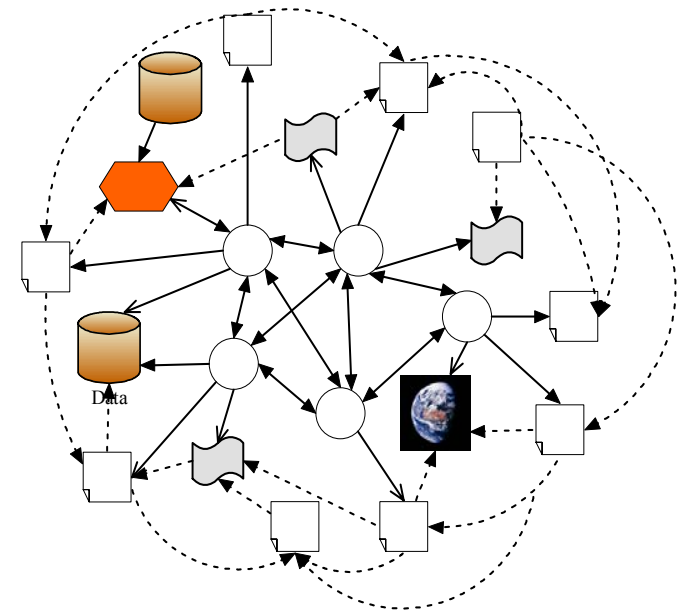
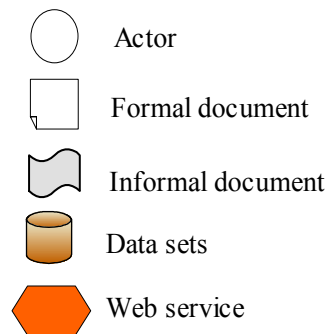


Integrating Social networks and bibliographic networks



Disconnected networks:

- formal publication network
- social network (actors)



Hybrid network

- documents (formal and informal)
- data
- services
- actors

Solution: Information Network Overlay

