

Motivation, inspiration and innovation from frustration

Herbert Van de Sompel
Research Library
Los Alamos National Laboratory, USA

Outline

- Exploration of ideas from 'Roadblocks' white paper
NSF Post Digital Library Futures Workshop
Cape Cod, June 15-17 2003
 - Brainstorm about DL research agenda for the next 10 years
 - Motivated by the Atkins report on Cyberinfrastructure: science and engineering research/education evolving towards global, network-based collaboration across institutions, disciplines, distance, time (cf. Earth science, Space Physics & Aeronomy Research, Ecology, High energy physics, Bio-informatics, ...)
 - Similar evolutions in humanities, social sciences, arts, ...

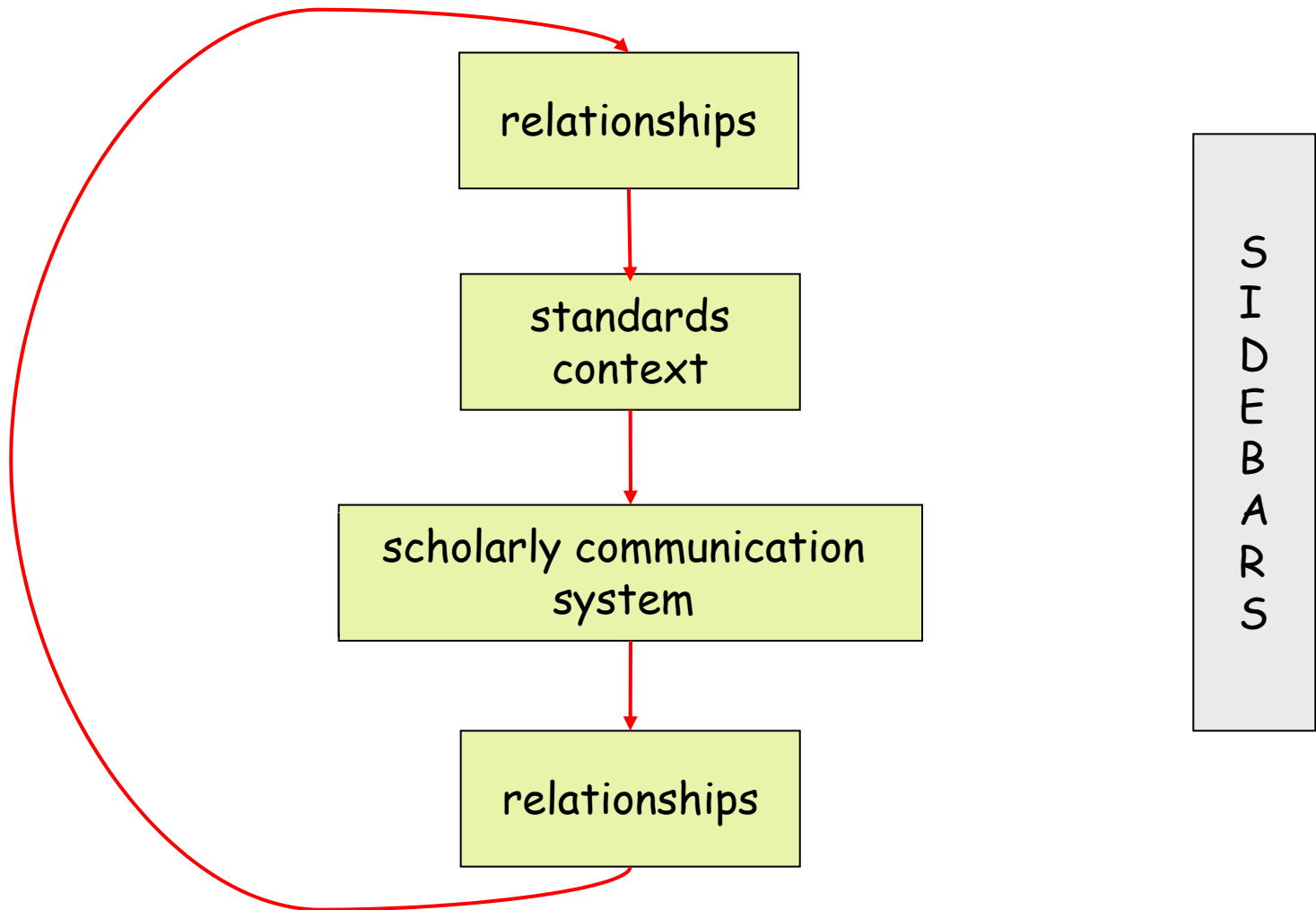
http://www.sis.pitt.edu/~dlwksshop/paper_sompel.html

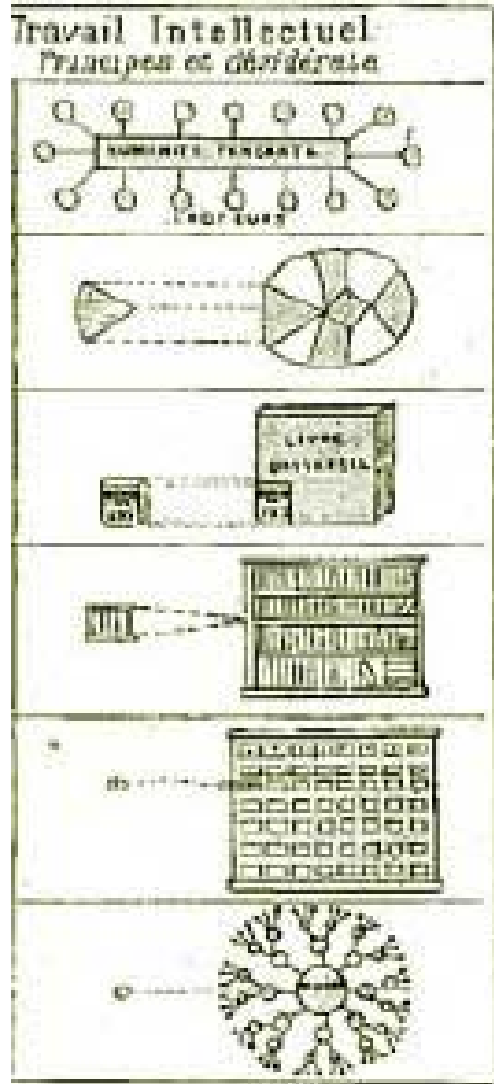
Outline

Roadblocks:

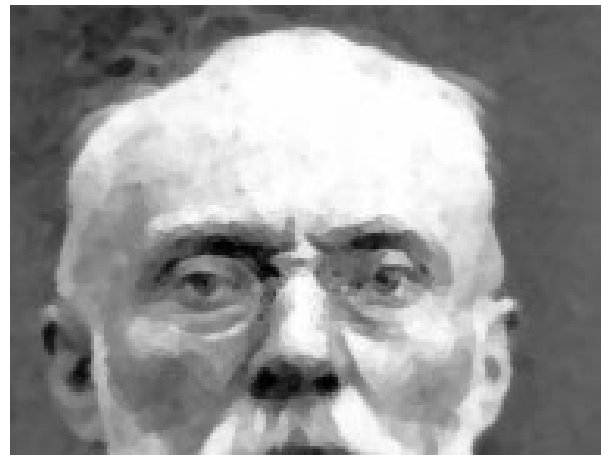
- Try and identify 'roadblocks' that preclude realizing the Cyberinfrastructure vision.
- All identified issues:
 - Seriously hinder whichever effort aimed at the creation of an integrated scholarly knowledge environment
 - Pose enormous research and organizational challenges at technical, legal, economical and social level
- Context of this presentation: information technology, networked libraries, publishers, scholarly endeavors including research, communication, education, ...

Outline





Paul Otlet



7040		TERZOKIS		DROSSOS		3		151028		201157		1	
No. copies		Title		Author		Year		Date		Date		Date	
27	4			1495		1	250	131	76	1	204	40	
28				1051	105					9	1850	40	
30	4			1071		1	77	94	38			40	
31	76			1344				118	57			40	
32	76			2558			147	225	239			40	
33	5			1880				147	103			40	
34	5			1861			231	167	128			40	
35	5			1630				143	96			40	
36	5			1854		1	224	162	128			40	
37				1680				147	103			40	
38	5			1071		1	210	127	134			40	

http://www.boxesandarrows.com/archives/forgotten_forefather_paul_otlet.php

Everything is Linked

How Everything Is Connected to
Everything Else and What It Means for
Business, Science, and Everyday Life

Linked



"Linked could alter the way we think about all of the
networks that affect our lives." —*The New York Times*

Albert-László Barabási
With a New Afterword

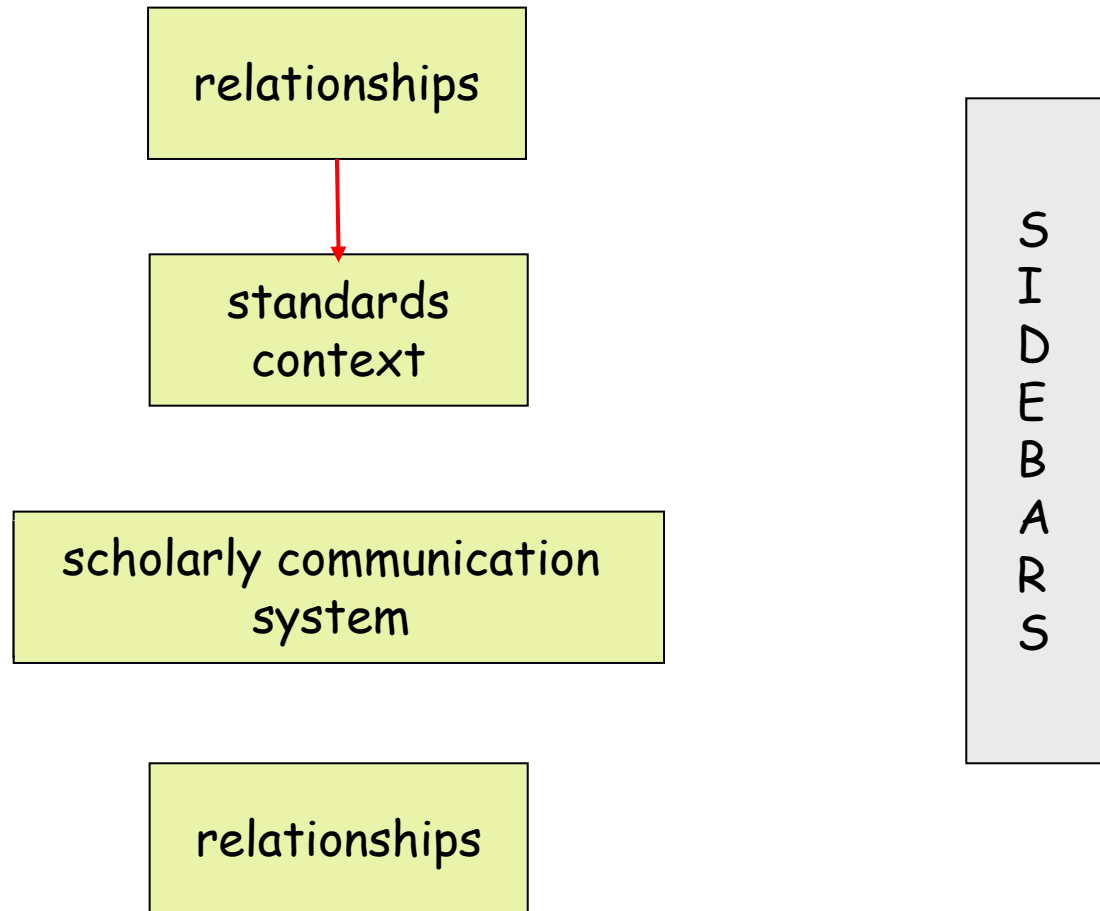
- But not so in the scholarly communication system: assets that are inherently related are not technically linked
- Scholarly system does not record an unambiguous trace of value chain of a scholarly asset
- Relationships known at the moment an assets goes through a step in a value chain are lost the second thereafter
- The dynamics of scholarship are lost. Such information can be crucial for the development of new knowledge.

<http://www.amazon.com/exec/obidos/ASIN/0452284392/>

Everything is Linked

- Early steps towards enabling technologies:
 - RDF to express semantic/knowledge relationships
 - DOI/CrossRef to hardwire 'cites' relationships in journal articles
 - OpenURL to dynamically and context-sensitively express metadata-based relationships

Outline



Standards/Interoperability context

- Standards are the glue that holds the networked information environment together.
- Standards are crucial to facilitate the emergence of improved and integrated services across networked libraries, across nodes in the scholarly communication system.
- As the information environment becomes more complex, and as we move towards new levels of services, we will need more, not less standards.

Standards/Interoperability context

- Standardization efforts/bodies in our community are seriously challenged:
 - Many standards defined outside our community.
 - Lack of impact on major standardization bodies of the networked world (W3C, IETF, IANA, ...)

Standards/Interoperability context: info URI scheme

http://www.info-uri.info/registry/OAIHandler?verb=ListRecords&metadataPrefix=oai_dc - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print W PDF W Add Selection

Address http://www.info-uri.info Go Google bill moen standards Links Net Snippets Add Selection

"info" URI Scheme

[About "info"](#) | [Internet-Draft](#) | [Registry ID](#) | [Registry Entries](#) | [Registry Policy](#) | [Register Namespace](#) | [Feedback](#)

RSS

Registry Entries

info:bibcode/	Namespace of Astrophysics Data System bibcodes
info:ddbj-embl-genbank/	Namespace of identifiers for sequence records in DDBJ/EMBL/GenBank
info:doi/	Namespace of Digital Object Identifiers
info:lccn/	Namespace of Library of Congress Control Numbers
info:oclcnum/	Namespace of OCLC Worldcat Control Numbers
info:ofi/	Namespace of Registry Identifiers used by the NISO OpenURL Framework Registry
info:pmid/	Namespace of identifiers of PubMed records
info:refseq/	Namespace of identifiers for RefSeq reference sequence record
info:sici/	Namespace of Serial Item and Contribution Identifiers
info:sid/	Namespace of Source Identifiers used in the NISO OpenURL Framework
info:srw/	Search/Retrieve Web Service

Start VALA2004 CSTB Pu... Re: OAI ... http://... C:\herb... VALA - N... Firework... Internet 3:08 PM

- Bridge between 'legacy' identifiers for information assets and Web infrastructure
- Essential for OpenURL Framework, Web services
- Useful for Semantic Web
- But ...

http://info-uri.info

Standards/Interoperability context

- Standardization efforts/bodies in our community are seriously challenged:
 - Many standards defined outside our community.
 - Lack of impact on major standardization bodies of the networked world (W3C, IETF, IANA, ...)
 - Problems to interconnect within and amongst related efforts in our community: digital library, grid computing, e-learning, library automation, ...
 - Operational models/processes not adequately adapted to the realities of the networked world (cf. patent challenges OpenURL, MetaSearch)
 - Funding for standardization efforts and related infrastructure is very hard to find (cf. OAI, CIMI, info URI Registry, OpenURL Registry, ...)

Standards/Interoperability context: Bill Moen

Dr. William E. Moen: Luminary Lectures @ Your Library - Microsoft Internet Explorer

Address: <http://www.loc.gov/rr/program/lectures/moen.html>

The Library of Congress >> Virtual Programs & Services >> Luminary Lectures

Luminary Lectures

@ your library™

Dr. William E. Moen

No Longer Under Our Control: The Nature and Role of Standards in the 21st Century Library

Wednesday, December 3rd, 10:30-12:00 noon
@ the West Dining Room (James Madison Building, [Sixth Floor](#))

ABOUT THE CYBERCAST:

[View Cybercast](#) (1 hour, 28 minutes - requires Real Player to view).

No registration or password required to access this cybercast lecture. The video of the lecture will be presented in RealPlayer format. To view it, you must have the Real Player installed and at least a 28 K-bps (kilobits per second) Internet connection for your computer. The RealPlayer software may be downloaded, free of charge, from the [RealNetwork Web site](#).

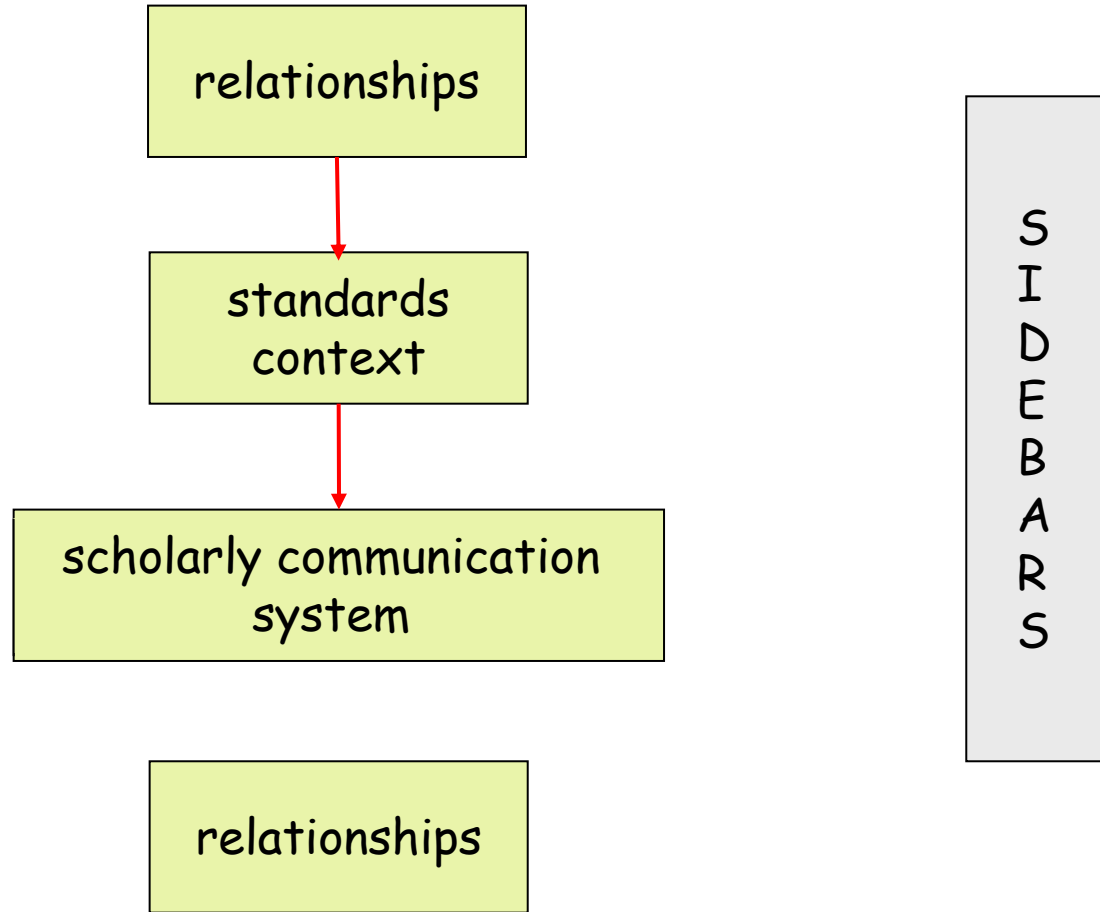
ABOUT THE LECTURE:

A MS Word version of the [lecture script](#) is available.

Libraries have prospered because they developed and adopted of a wide range of standards, guidelines, and community practices to serve their needs. Broadly speaking, standards reflect vital community agreements on problems and their solution. The networked environment, however, is changing the context and use of standards in libraries. Librarians no longer create and define the terms, concepts, standards, and technologies that drive library services and practices.

<http://www.loc.gov/rr/program/lectures/moen.html>

Outline



A few bold statements about the scholarly communication system

- The current electronic scholarly communication system is a mere scanned copy of the paper-based system.
- The networked environment begs for another scholarly communication system to emerge.
- We really need to start creating that other system.

Absorbing and distilling a disruptive new technology

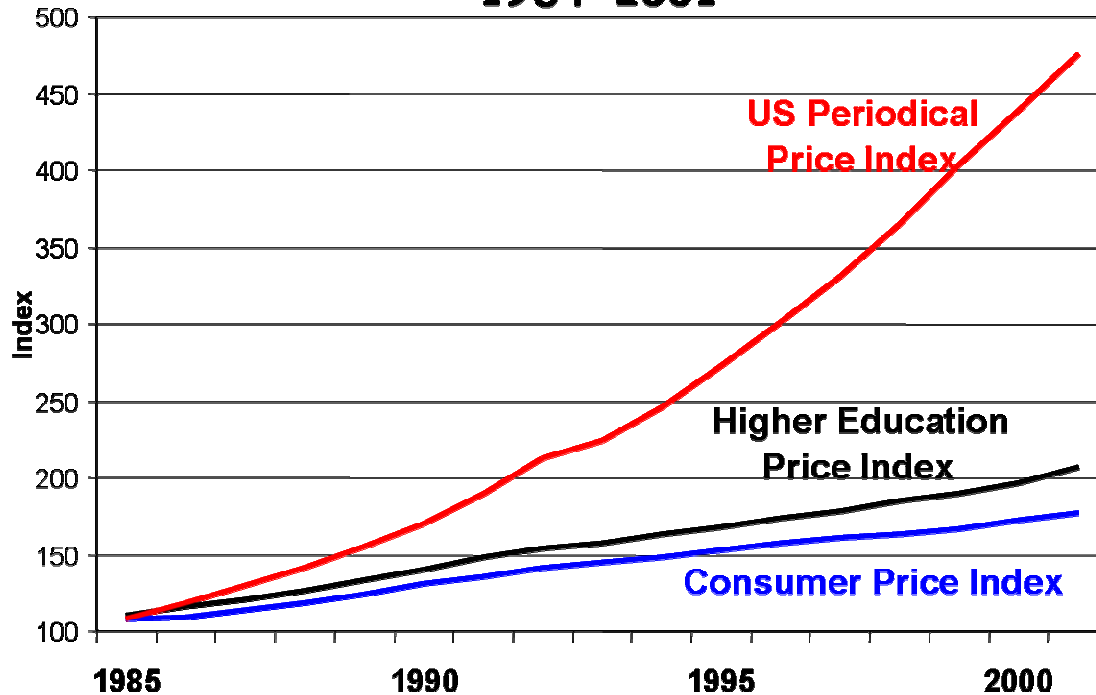
- Three stages of technology change
 - **Modernization** → ejournals
 - Do **things we couldn't do before** - e.g., linking, personalization, on-line peer-review/commentary, annotation, ...
 - **Transformation** : Real gains in the application of the new technology. Requires:
 - social-organizational change
 - a systems perspective

What is driving the Need for Change?

- The Usual:
 - Serials crisis
 - Permission crisis
- The New ~ networked environment
 - Repository model
 - E-science
 - Rights framework

The Usual: serials crisis

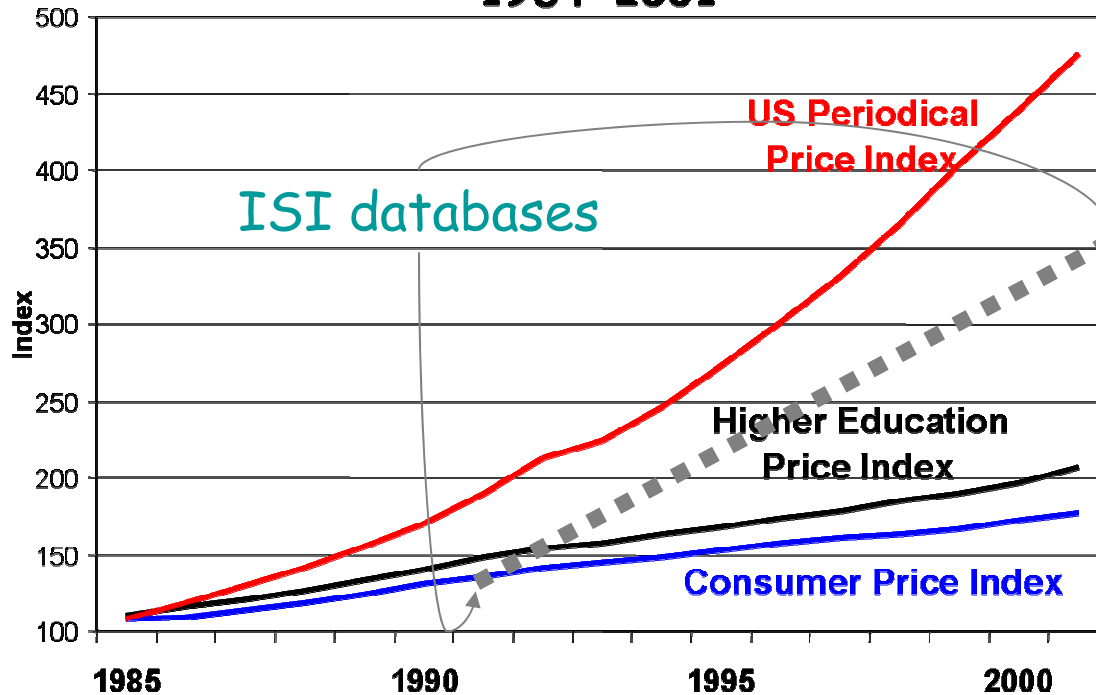
**Periodical Price Increases in Comparison with
Common Inflation Indexes:
1984-2001**



http://www.ala.org/Content/NavigationMenu/Products_and_Publications/Periodicals/American_Libraries/Selected_articles/charta.htm

The Usual: serials crisis

Periodical Price Increases in Comparison with Common Inflation Indexes: 1984-2001



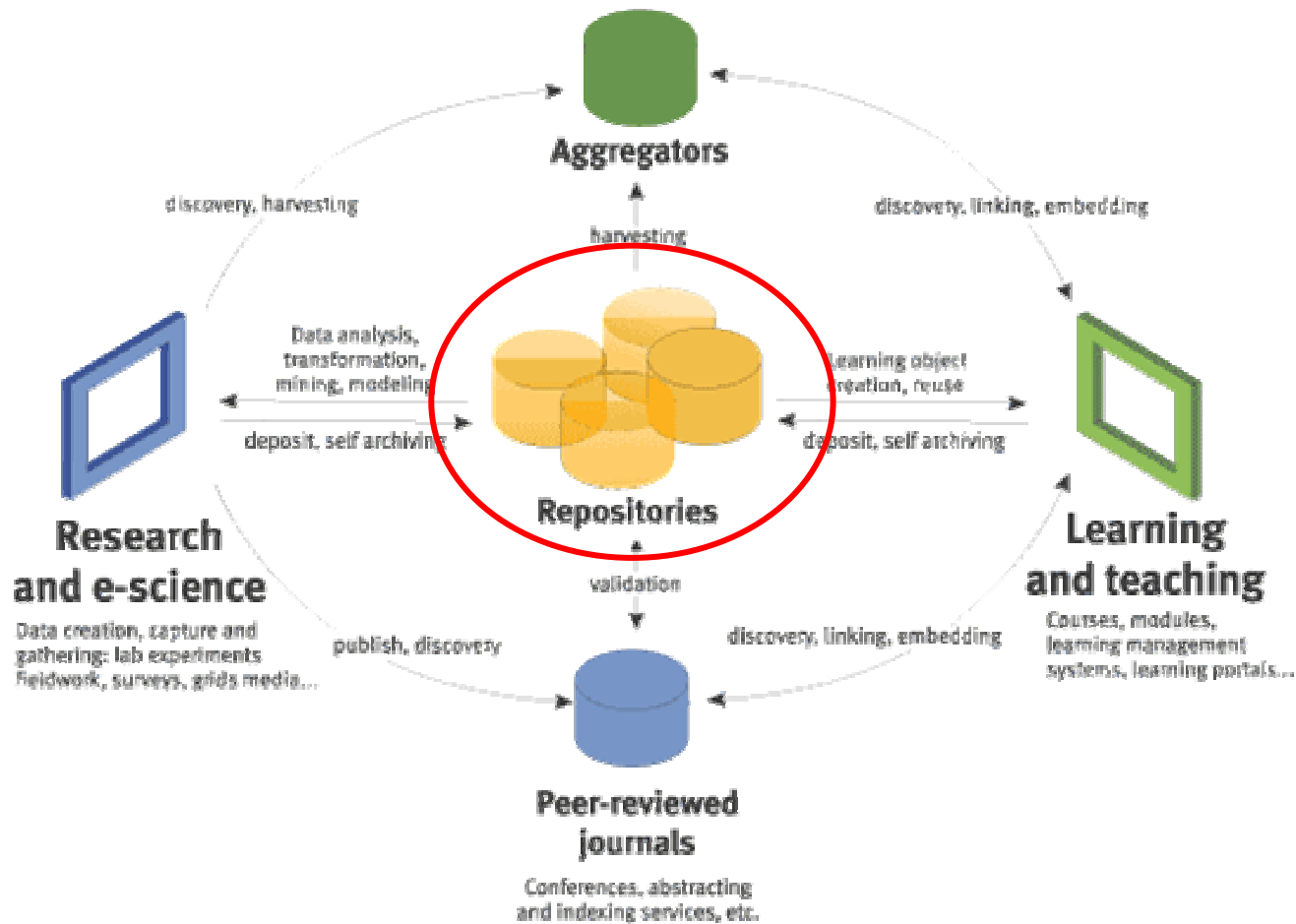
- Competition-less market
- Related to policies in academia re promotion, tenure, funding based on ISI metrics?

Price for ISI databases as paid by LANL 1991-2002

The Usual: permission crisis

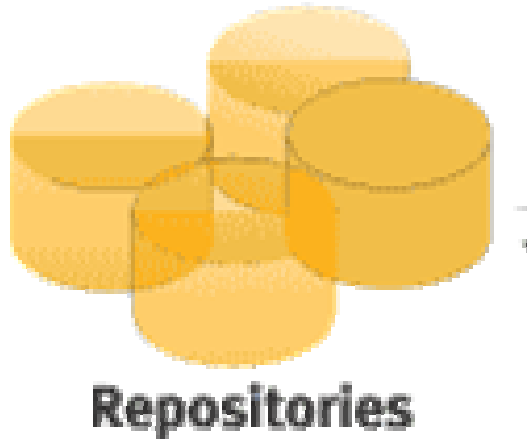
- Technological barriers limiting libraries' use of journals
 - Control of the IT that adds value to content
- Legal barriers - copyright & licensing contracts
 - Some rights generously granted by ...: *"Authors may distribute or transmit their own paper to not more than 50 colleagues. Authors may post the title, abstract (no other text), tables, and figures from their own papers on their own Web sites."*

The New: repository model



"Pattern Recognition: The 2003 OCLC Environmental Scan"
<http://www.oclc.org/membership/escan/toc.htm>

The New: repository model



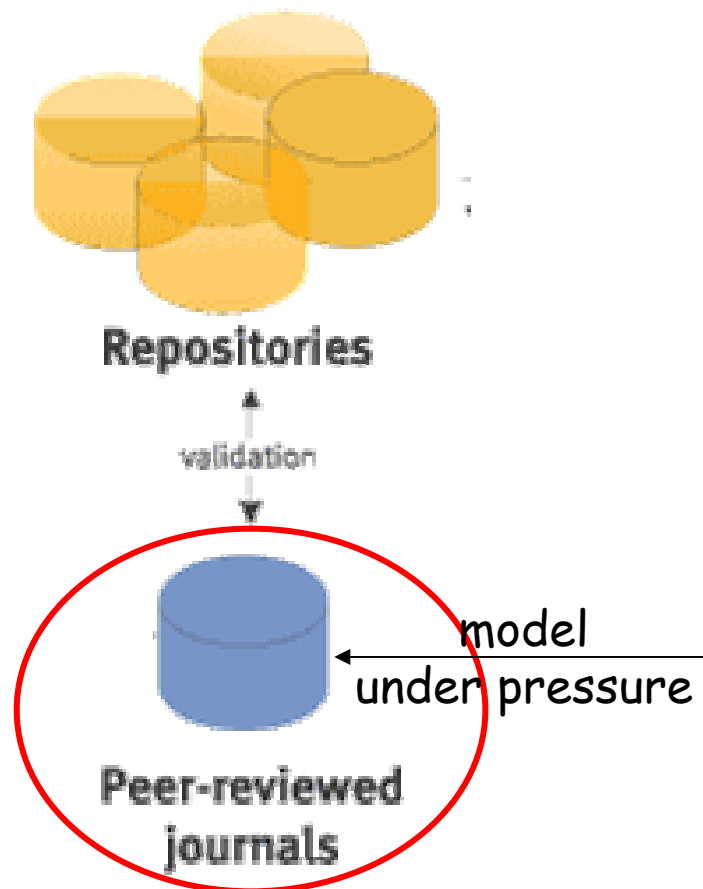
- Different repository types: IR, Grid repositories, digitized materials, ... **no interoperable model**
- Different media types, document models, ... **no interoperable model**
- Different services ... **no interoperable model**

The New: repository model



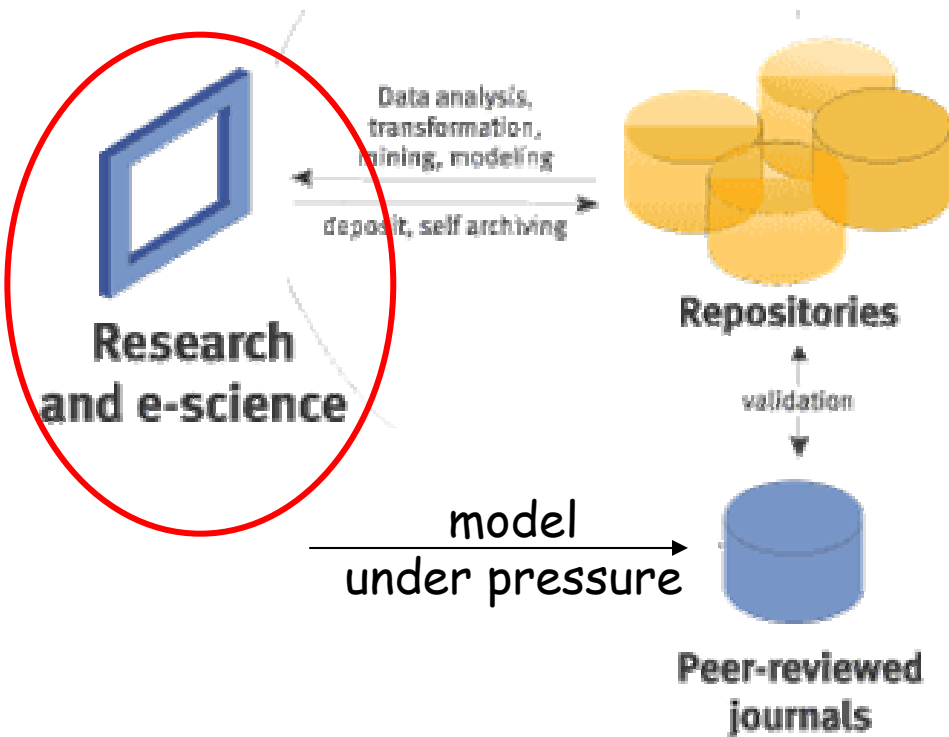
- Different levels of 'vetting': no review, internal review, curatorial decisions, peer-review, ...
- Focus on raw materials:
 - cf. arXiv ~ APS evolved towards integration
 - start of scholarly value chain **S**

The New: repository model



- Journal system is just one possible, vertically integrated value chain
- In a networked world, the functions it performs can/will be handled in a deconstructed/distributed manner:
 - registration in repository
 - validation by different nodes/parties
 - archiving by different nodes/parties
 - awareness by different nodes/parties

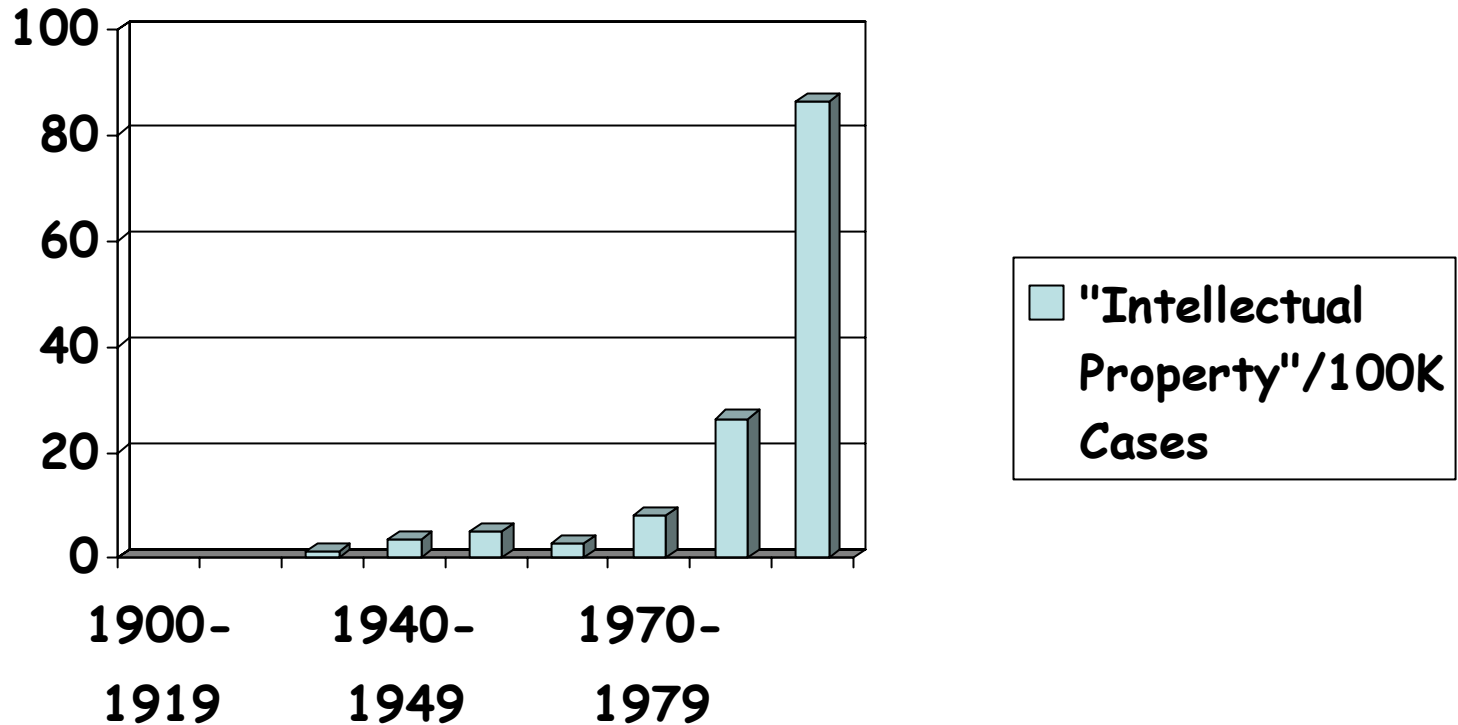
The New: e-science



- Results of data analysis, transformation, mining, modeling need to be validated too
- These become 'publications' in their own right, the start of scholarly communication value chains
- Journal system not well placed for this:
 - Lacks review mechanisms
 - Can't handle the nature of 'new' publications
 - Bad fit with nature of e-science: speed, dynamic, collaboration

The New: rights framework

occurrences of the phrase "intellectual property" per 100,000
U.S. Federal Cases



"Professor Hank Greely" Cited in Lessig, L. The future of ideas: the fate of the commons in a connected world. NY, Random House, 2001. P. 294.

The New: rights framework

- Urgent need for an environment in which scholarly assets behave in a manner that matches the “gift exchange” spirit of scholarship. James Boyle: Think about what we lose by sticking with the current paradigm!
- In this context:
 - complexities built into systems to accommodate existing rights framework
 - enormous constraints on ability to use scholarly assets: process to extract knowledge, attach knowledge, mine, evolve, build upon: **robots are the next generation readers**

Robots as readers

PMC Open Archives (OAI) Service

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History

Address <http://www.pubmedcentral.gov/about/oai.html> Go Google vava 2004 openurl

PubMed Central PMC OAI Service

About PMC Journal List Search Utilities

PMC Utilities

- ▶ OAI Service
- ▶ Open Access List
- ▶ pmc-1 XML to HTML
- ▶ E-utilities

OAI Access to PubMed Central Records

The **PubMed Central OAI service (PMC-OAI)** provides access to metadata of all items in the PubMed Central (PMC) archive, as well as to the full text of a subset of these items.

PMC-OAI is an implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a standard for retrieving metadata from digital document repositories. Visit the [Open Archives Initiative site](#) for more information about the protocol and other activities of the OAI group.

PMC-OAI supports **OAI-PMH version 2.0**. Compatibility with earlier versions of the protocol is not guaranteed.

Using the PMC-OAI Service

The **base URL** for the service is <http://www.pubmedcentral.gov/oai/oai.cgi>.

High-Volume Retrievals

If you are using a script that makes more than 100 requests of any kind, please run it outside of the PMC system's peak hours.

Do not make more than one request every 3 seconds, even at off-peak times.

Peak hours are Monday to Friday, 5:00 AM to 9:00 PM, U.S. Eastern time.

Access to Full Text

- Full content
OAI-PMH
harvestable from:
- PubMed Central
 - BioMed Central
 - PLoS (soon)





<http://www.pubmedcentral.gov/about/oai.html>

Towards an alternative ecology?

- Another rights framework
- Pathways model
- Libraries in a new ecology
- Barter trade model

You're probably familiar with the phrase, "All rights reserved," and the little (c) that goes along with it. Creative Commons wants to help copyright holders send a different message: "Some rights reserved."

For example, if you don't mind people copying and distributing your online image so long as they give you credit, we'll have a license that helps you say so. If you want people to copy your band's MP3 but don't want them to profit off it without your permission, use one of our licenses to express that preference. Our licensing tools will even help you mix and match such preferences from a menu of options:

-  **BY: Attribution.** Permit others to copy, distribute, display, and perform the work and derivative works based upon it only if they give you credit.
-  **Noncommercial.** Permit others to copy, distribute, display, and perform the work and derivative works based upon it only for noncommercial purposes.
-  **No Derivative Works.** Permit others to copy, distribute, display and perform only verbatim copies of the work, not derivative works based upon it.
-  **Copyleft.** Permit others to distribute derivative works only under a license identical to the license that governs your work.

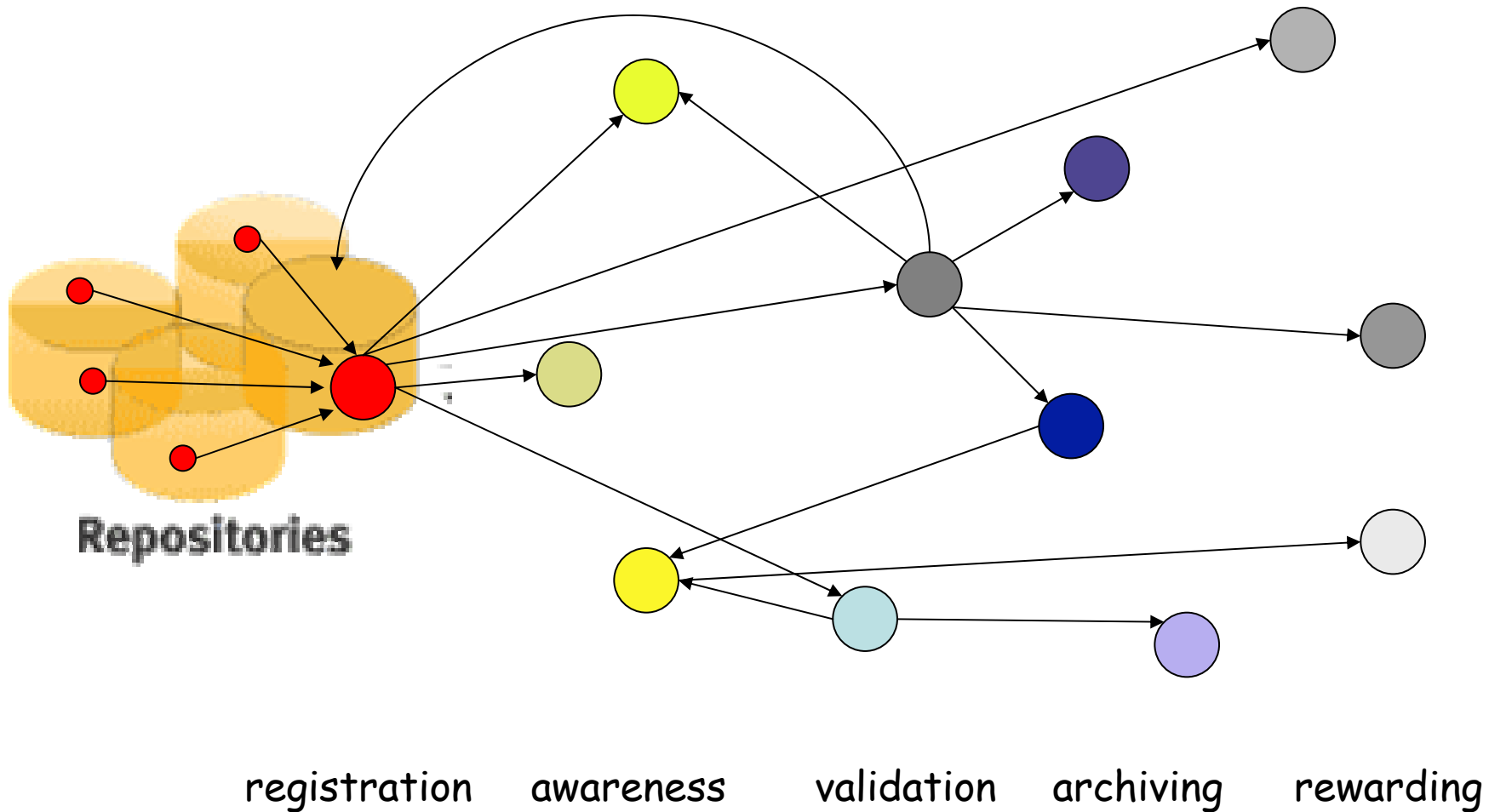
OAI-rights effort



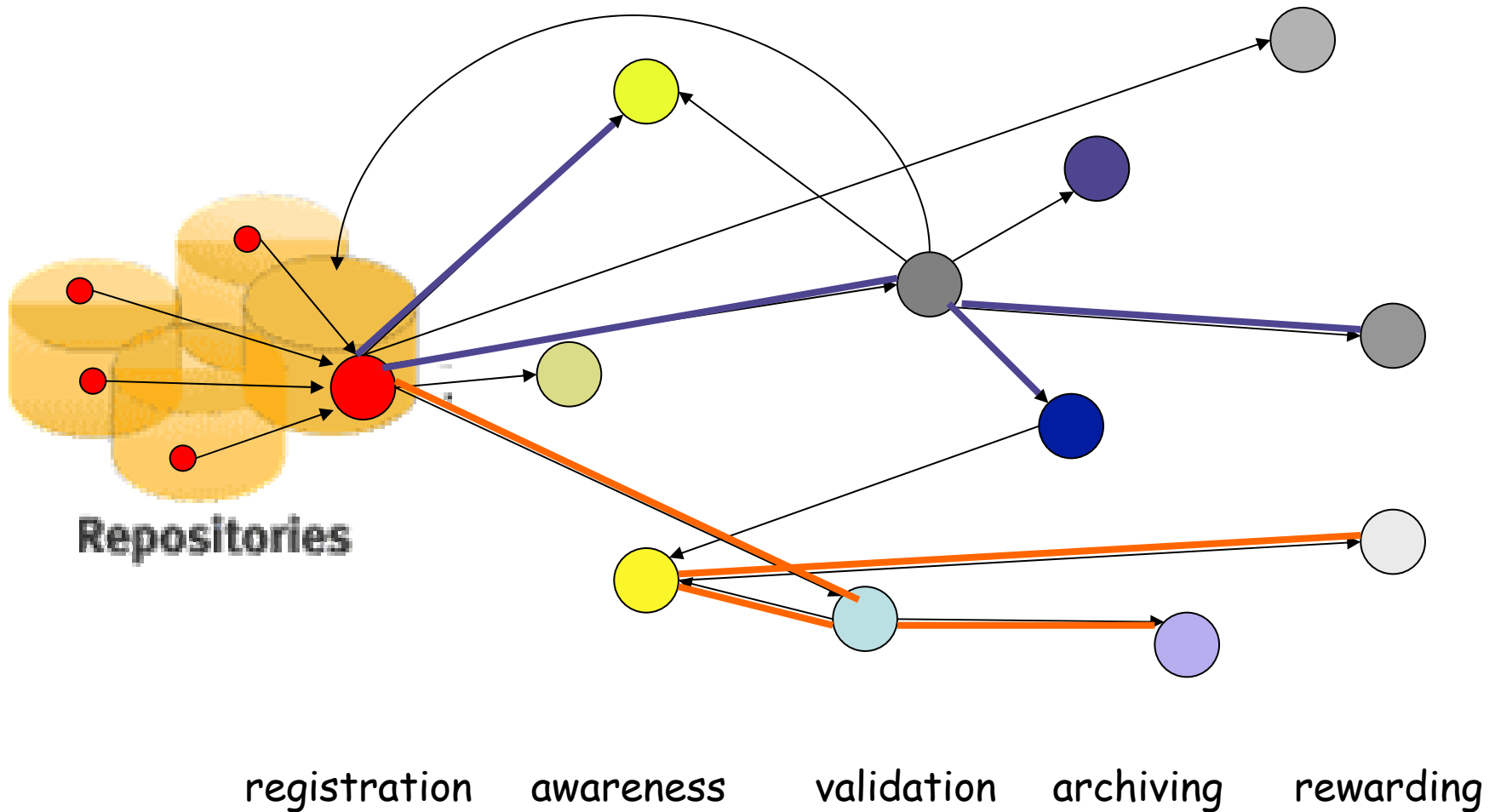
- Convey rights about metadata
 - Convey rights about resource
- in the OAI-PMH framework
- Special attention for Creative Commons licenses

<http://www.openarchives.org>

Pathways model



Pathways model



Pathways model

- Many nodes perform functions of scholarly communication:
 - indexing, searching, recommendation, linking, data-mining, visualization, ... nodes
 - annotation, certification, metric-collecting, rewarding, ... nodes
 - archiving, normalization/transformation, ... nodes

Libraries in a new ecology?

"They are only there to pay the bills"

- Libraries need to rethink their position. They can push towards establishing an alternative scholarly communication ecology:
 - Content-nodes
 - Service-nodes

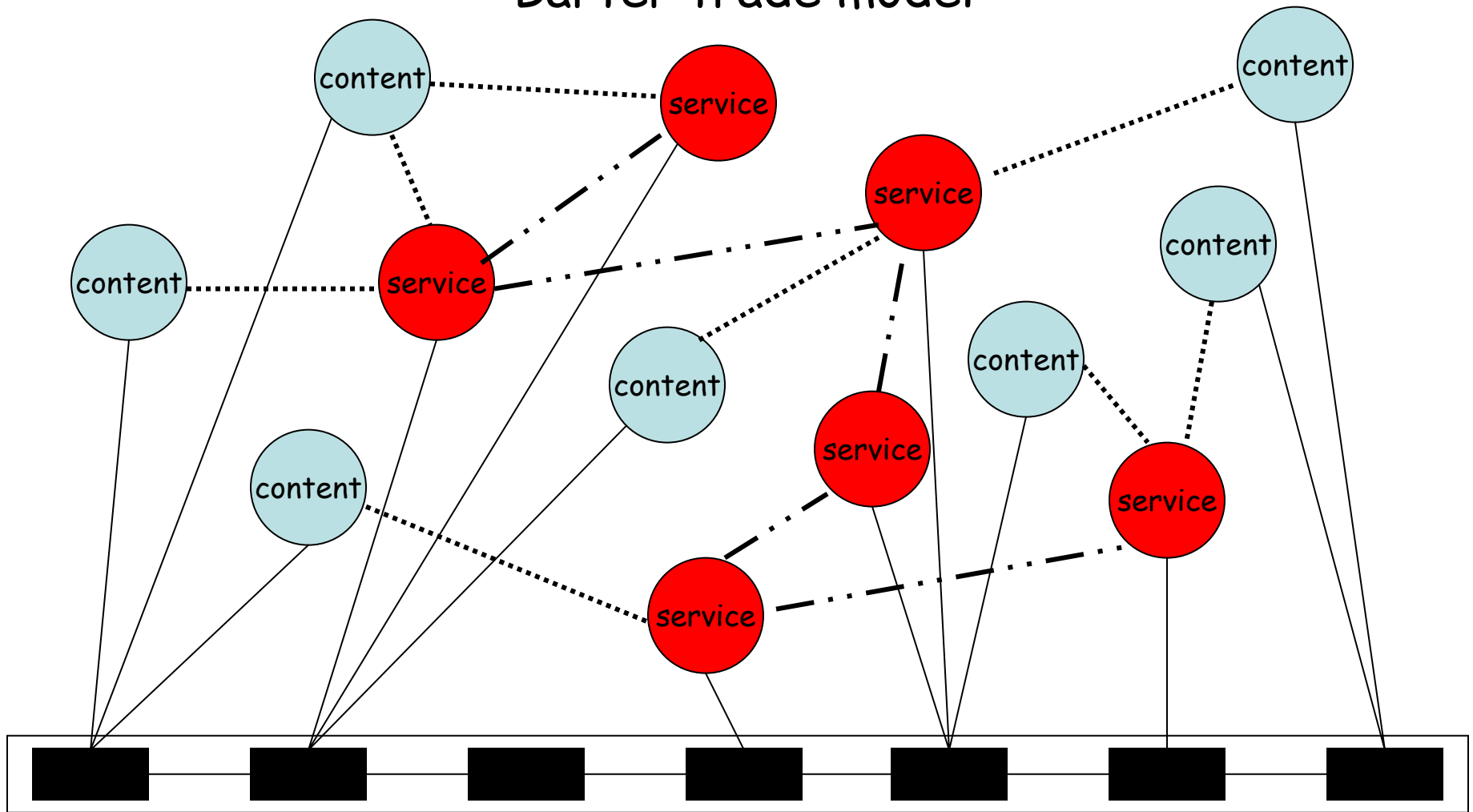
Ongoing: Libraries at the start of a new value chain

- Institutional Repository network:
 - Libraries become content-nodes, capturing the intellectual output of their parent institutions and “exposing” it.
 - Vision: A network of federated repositories that makes available the collective intellectual output of faculty and researchers of the world's research institutions

Next: Libraries as service nodes in a new value chain?

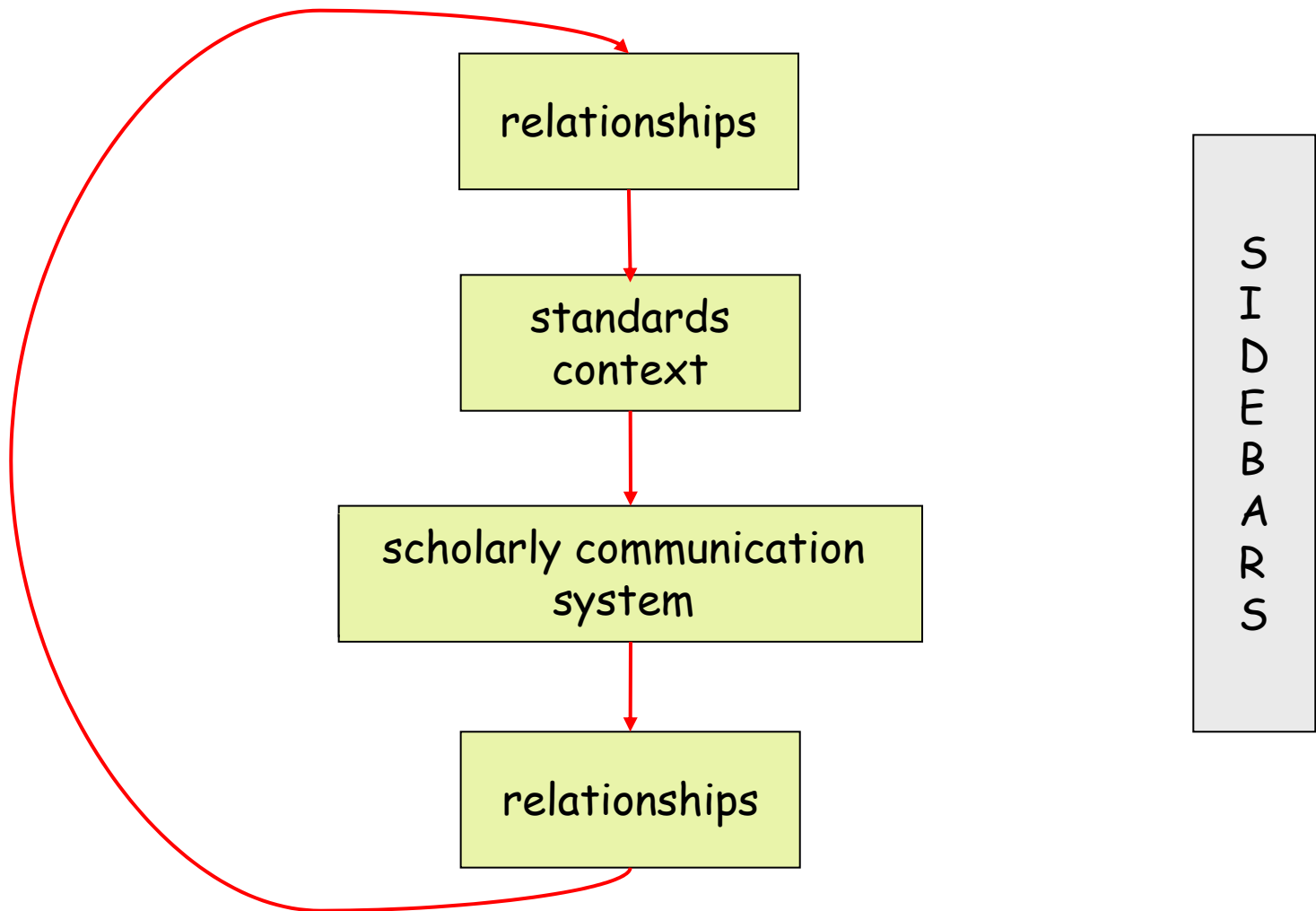
- Redundancy in services provided by libraries, following the "paper-based tradition". Seems non-optimal in networked environment.
- Complexity of IT services increases. Many libraries will eventually be forced out of the "service" business (read: out of the "all services" business).
- Trend: Specialization & distributed collaboration: OAI-PMH model, Seti, LoCKSS, QuestionPoint
- Vision: A federation of networked services (web services) - in which Libraries take on specific service tasks - that turns into a new global scholarly value chain

Barter trade model



content/service Clearing House

Outline



Relationships

- Need to record/expose scholarly dynamics, relationships and interactions between assets: many arcs, many types of relationships interconnect assets and actors.

Conclusion

The networked environment might allow us to evolve towards another scholarly communication environment:

- That reflects the gift exchange spirit of scholarship
- That is more flexible because modular
- That allows for nodes to compete
- That is 'really' networked
- That records relationships within the infrastructure

Conclusion

- OpenURL and OAI-PMH are the results of frustration with the state of affairs.
- Be frustrated about the current state of affairs.
- Use the frustration as motivation and contribute to change.
- **Become the next keynote speaker.**