Bridging the Past and Future: Scholarly Communication in the 21st Century

Carl Lagoze
Cornell Information Science
lagoze@cs.cornell.edu

Abstract

The fabric of scholarly communication is undergoing radical change. Economic, legal, and technical tensions are straining traditional publisher and journal models. New technologies provide unexplored opportunities for new scholarly communication models. At Cornell Information Science we are exploring this problem from the perspective of social and technical interaction. Our work therefore involves investigation of new infrastructure for scholarly communication, in particular the extension of OAI-PMH to a p2p framework, and research into hybrid social/technical networks that arise in the context of new technical infrastructure.

1 Introduction

The traditional model of scholarly publishing – publication through peer-reviewed journals – is under enormous pressure. Publishers and libraries are entwined in what is referred to as the "serials crisis" [23], where rising subscription rates, flat library budgets, and resulting subscription cancellations threaten the very existence of journals and research library collections. In addition, a growing number of scholars, especially those in rapidly changing scientific fields, have become increasing accustomed to rapid dissemination via the web and are frustrated with the delays of journal publication [41]. Finally, scholars have growing resistance to copyright transfer policies that they perceive as vehicles for giving away their intellectual capital [11].

The ubiquity of the web and internet has given rise to a number of developments – collectively known as scholarly electronic publishing [9] – that are effecting radical change to the scholarly publishing milieu. These include electronic publishing by existing commercial and learned society publishers [13], growing popularity of e-Print archives that provide discipline-centered repositories for rapid exchange of scholarly results [38], rapid growth of so-called institutional repositories [26] whereby institutions and their libraries provide the infrastructure for self-publishing, and simple self-publishing whereby scholars maintain active web sites for research projects.

These developments have already had a transformative effect on the way scholars disseminate information and the manner in which libraries manage that information. Nevertheless, they can be characterized as *incremental* developments, borrowing the terminology of Kling and McKim [30].

These developments are incremental in that they maintain the tradition of document as text, albeit in digital form. While the importance of text should certainly not be minimized, there is substantial demand from both the scientific [53] and humanities [48] communities for scholarly artifacts that interweave text, static data, simulations, video, and other genre of materials. The opportunities in the digital domain for extending documents beyond text have been explored by Buckland [12] and our own work with Fedora [3] provides the infrastructure for such new documents.

These developments are incremental because they focus on artifacts, rather than the process that produces those artifacts and the combination of that process with the artifacts in a coherent network. As Kling, McKim, and King note [31], intellectual activity arises from what they call "scholarly communication forums" that intermix people, organizations, data, messages, programs, and documents in rich "socio-technical interaction networks". In the same sense that amazon.com has elevated numerous entities such as reviews, lists, and cross-references to a first-class status (alongside the products they sell), we maintain that next-generation scholarly communication systems should reify as "documents" not only results, but the signal-rich communication that up-to-now has been lost in email archives or discussion lists and the relationships of those interactions with related scholarly documents.

Finally, these developments are incremental because they largely maintain organizational structures inherited from print publishing. For example, arXiv, arguably the most influential of the new publishing vehicles, essentially maps a pre-digital tradition of sharing preprints among physics colleagues onto web space [30]. The interaction of communication technology and organizational structure has been relevant since the invention of the book [25, 40]. There can be little doubt that the organizations that support scholarly communication will radically change in the years to come in tandem with radical changes in publishing technology.

In a variety of projects at Cornell Information Science, we are investigating infrastructure and theory underlying present and promising changes in scholarly communication. This work effectively "bridges the past and the future", in that it attempts

to resurrect the interpersonal/interactive foundations of scholarly communication, which have been lost due to the physical constraints of paper-based journal publishing, while accommodating the "beyond text" needs of modern scholars. This paper will briefly outline two aspects of this work that are planned or are in process and hopefully stimulate discussion in this rapidly changing area.

2 P2P and the Open Archives Initiative

p2p systems are the subject of considerable interest in the research community [17] because they can be highly adaptive, self-organizing, and fault tolerant. We are interesting in providing technical infrastructure that will enable the deployment of p2p in the context of scholarly communication. Our primary interest is to understand how p2p architecture facilitates the *scholarly value chain* whereby ideas flow from proposals to initial results to preprints to peer-reviewed publications. Can p2p enable a richer, more democratic, and ultimately more productive scholarly value chain?

There has been limited work on p2p systems within the digital library domain or related to scholarly publishing. LOCKSS implements p2p replication for the preservation of documents [37, 46]. Its main contribution is the use of novel techniques to ensure the survivability of document copies even in the face of hostile attacks. The results of the Edutella project [39] are more closely related to scholarly publishing concerns. Edutella exposes RDF metadata [6] via OAI-PMH [7] in a P2P framework and provides a query capability over that metadata. Query routing is done via hub peers that rely on the semantic structure of the RDF metadata. Finally, the W3C sponsored Annotea project [27] and in particular its extensions in mies [8] addresses issues related to p2p sharing of annotations and comments about digital documents.

Our investigations of P2P within the context of scholarly communication build on the notion of a *metadata refinement network* that enables the creation of *documen value chains*. Such a value chain exists when distributed nodes in the network iterate an original resource *r1* by creating other types of manifestations *r2*, *r3*, …, *rn* of it, by, for example, expressing a quality statement about it, annotating it, digitally preserving it, etc. Presuming that metadata *m1*, *m2*, *m3*, …, *mn* is associated with each such iteration, we are interested in investigating how that distributed metadata can indicate its relationship to the original resource *r1* and/or to other iterated versions *r2*, *r3*, …, *rn* of that resource.

To create the infrastructure for the evolution and dissemination of metadata to support such value chains, we are extending OAI-PMH to a p2p context. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [33] is widely recognized as the predominant architecture for interoperability among distributed information sources. The protocol is used internationally for the exchange of structured (meta)data in a variety of contexts including digital libraries, museums, ePrint repositories, research projects and corporate intranets. OAI-PMH is quite simple, with six verbs that allow known item retrieval, range retrieval based on simple criteria (for example, date of modification), and retrieval of basic repository information. The simplicity of the protocol is key to its widespread use and utility and Clifford Lynch, a recognized leader in the information community has described the OAI-PMH as "a vital component of the digital information infrastructure" [36].

When the protocol was originally created, the primary usage context was a simple bipartite model, illustrated in Figure 1. In this model, users of the protocol are divided into two classes: *data providers* make structured (meta)data available via the protocol, and *service providers* use protocol requests to harvest the data, post-process it, and refine it with the goal of developing services that add value to the metadata.

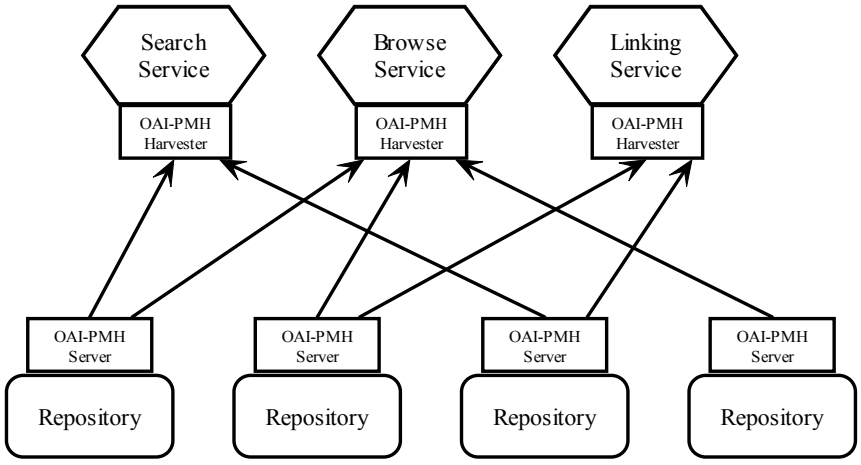

Figure 1 - Simple Data Provider, Service Provider Topology

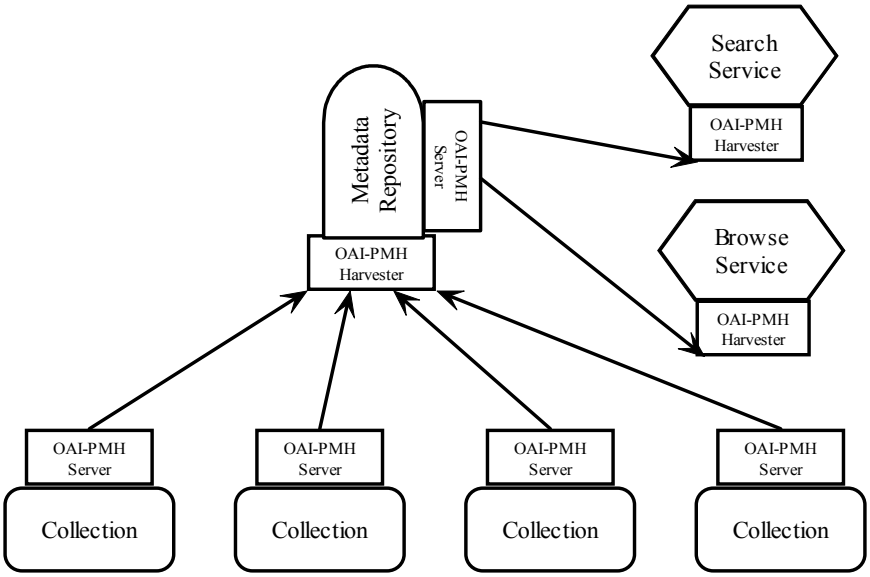

Figure 2 - NSDL Metadata Repository as OAI-PMH aggregator

This simple model on which the OAI-PMH was based provided a good basis for initial understanding of the utility of the protocol, leading to its rapid dissemination. However, use of the protocol quickly revealed another class of deployment. This is the notion of an *aggregator*, which in essence straddles the line between data provider and service provider. An aggregator consumes metadata via OAI-PMH from several sources (data providers) in the same manner as a standard service provider. In addition to storing that metadata, an aggregator may process it in several ways including normalizing it based on analysis of a large cross-section of aggregated metadata records and cross-walking between different formats, for example, MARC to Dublin Core [2]. The aggregator then exposes this metadata via OAI-PMH, effectively changing roles to a data provider.

The NSDL (National Science Digital Library) [54] metadata repository and its relationship with NSDL collections and services, illustrated in Figure 2, is an example of this aggregator model. This forms the basis of the NSDL architecture [32] in its current form at the time of writing this paper

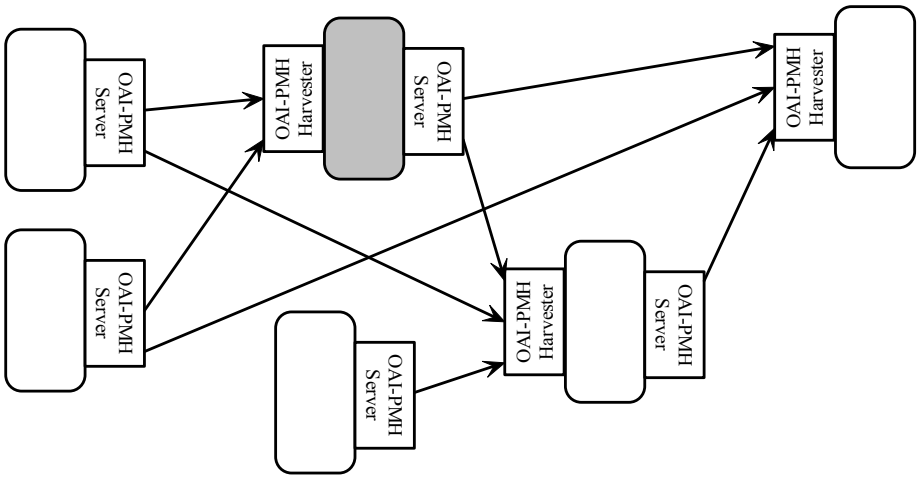

Figure 3 - OAI-PMH p2p network

(January 2004).

We suggest that this data-provider/aggregator/service-provider model is evolving towards a more intriguing topology of nodes participating in the exchange of structured data via OAI-PMH. For example, in our own work within NSDL we have found that harvesting among aggregators occurs; such as OAIster [5] harvesting from the NSDL metadata repository. We foresee further blurring of the lines between data providers, aggregators, and service providers and the evolution of a directed graph topology that corresponds to that of peer-to-peer (p2p) networks [42]. An example of this graph is shown in Figure 3, with nodes playing four roles based on their graph relationships: 1) *traditional data providers* with only outgoing arcs, 2) *traditional service providesr* with only ingoing arcs, 3) *pass-through aggregators* that consume and provide metadata but perform no transformations or enhancements, and 4) *value-added aggregators* that normalize and add value to consumed metadata before exposing it (shaded in grey in the figure).

The flow of data and metadata through such a network has a number of interesting characteristics that we are investigating and modeling.

2.1 Metadata document location

Similar to p2p document systems such as freeNet [14] and Gnutella [4], metadata (documents) in an OAI-PMH p2p network are distributed across the network and clients must be able to locate copies of specific metadata records. We can exploit the fact that metadata records in OAI-PMH are uniquely named and use efficient key-based location mechanisms such as distributed hash tables (DTHs) [45, 47] or other established distributed lookup protocols [49] to locate specific metadata records in p2p metadata network.

2.2 Provenance queries

Metadata records may propagate through the system and go through a number of refinement and/or translation phases as they move through *value-added aggregators*. This type of processing is typical of a number of aggregators run by existing libraries. At any given time *t* in a OAI-PMH p2p network, a set of metadata records *{m1, m2, m3, ...}* may exist that have a related metadata provenance chain. This provenance chain may be a simple and linear or may have multiple branches. In a related, and possible concurrent, scenario, an original resource *r1* evolves through a value chain in which iterated versions {*r2, r3*, … } of *r1* are being created. In this case, at any given time *t* in a OAI-PMH p2p network, a set of metadata records *{m1, m2, m3, ...}* may exist that have a related resource (not metadata) provenance chain. Clients in this OAI-PMH p2p scenario should be able to query the system for other metadata records of related provenance.

The OAI-PMH already has an implementation guideline [34] for expressing provenance information in a harvested metadata record and we are investigating whether this specification is sufficiently expressive and how it can be automatically generated. We are also investigating the techniques for joining records by provenance chain in order to efficiently query the system in this manner.

2.3 Network harvesting

Finally, and most interesting, it would be useful to address OAI-PMH harvesting requests (*ListRecords*) to an OAI-PMH p2p network as a whole. For example, a client might specify a request to harvest all records in the network that have been made available or modified between dates *d1* and *d2*. The implementation of this would involve submitting a query to a single node on the OAI-PMH p2p network that then efficiently propagated the query across the network. Broadcast queries, such as that which take place in Gnutella [4] have proven to be inefficient for this task. More promising is to address this as a selectively propagated *range query,* which is a problem that has been studied by our colleagues in the database group at Cornell. They have developed an efficient and elegant data structure called P-Trees [15], which notionally distribute b-tree functionality across distributed peers and can be used to efficiently route queries among peers. We are investigating this data structure as an approach to implementing efficient network harvesting.

3 Document networks, social networks, and hybrid networks

The development of printed journals transformed scholarship from a closed group activity relying on verbal face-to-face communication among colleagues to a more democratic and accessible system, albeit one where the hierarchy of excellence prevailed. Print publications provided the basis for what Anselm Strauss [50] described as “social worlds” where scholars worldwide developed a sense of community through texts rather than personal interaction. We are investigating how the structure of these “social worlds” mutates as the publication technology that underlies them is transformed.

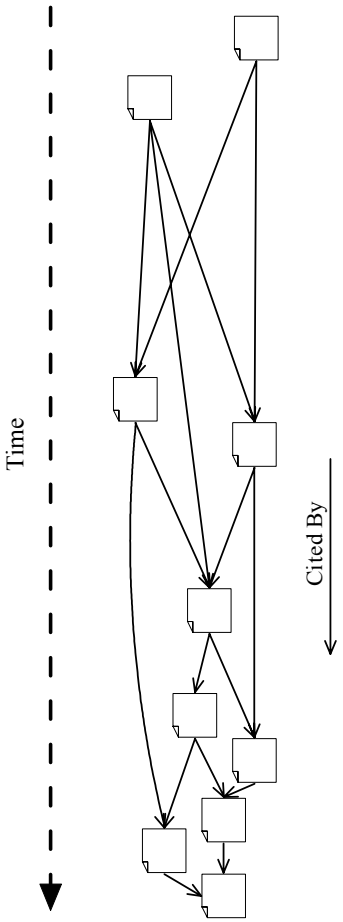

Figure 4. Scholarly Publishing Network

Our work integrates social science research methods to explore the interplay between information technologies and scholarly communication, and aims to design new tools to facilitate more dynamic and richer scholarly communication. Analysis of the use of these new tools will reveal the effects of these technologies and shed new light on our understanding of how scholars communicate with each other and how different communication models lead to different manifestations of scholar community.

3.1 A Conceptual Framework for Scholarly Communication

To clarify the interplay between information artifacts, the forming of a scholarly community, and its institutionalization, we propose a conceptual framework based on information flows. We argue that scholarly publishing and communication are all information exchange through information artifacts: either formal and materialized artifacts such as journal articles and conference proceedings, or informal and non-conclusive artifacts such as email discussions and informal meetings.

In order to facilitate the discussion, three classes of document can be identified: formal documents, in the form of officially published journal papers and conference proceedings; semi-formal documents, such as the ePrints in arXiv; and informal documents, such as email discussions, blogs, etc.

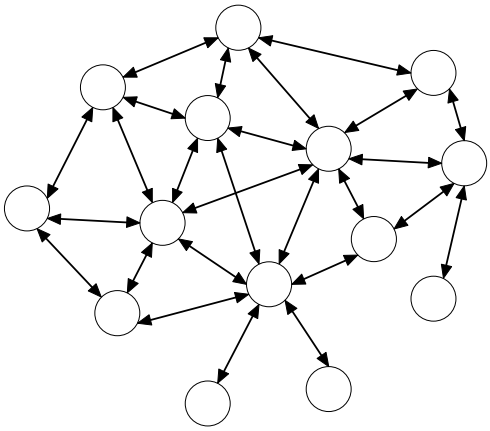

Figure 5. Researcher Social Network

The official and semi-formal publication systems form document networks (Figure 4). They are directed acyclic graphs with citation edges from newer papers to those already published. The development of electronic publishing exemplified by arXiv has collapsed the time dimension of this citation graph: one hundred years ago you might only be able to cite papers published in the past year; now you can cite contemporaneous papers. Evaluation of this citation graph reveals popularity of documents and suggests quality [20-22]. Substantial research has been done regarding why people cite each other, including quality, same field, availability, self-citation, and social citation, indicating that the system is far from perfect [10]. Furthermore, as previously discussed, a large amount of informal communication and the related scholarly artifacts were neither "materialized" nor captured nor analyzed, and were consequently excluded from quality and popularity measurements.

Coexisting with the document networks, there are social networks composed of authors and researchers. Leydesdorff [35] discussed a dual layered network, including the social network of authors and network of their reflexive communications. Cronin [16] also discussed citation as one dimension of scientific communication among texts. Social networks among scholars are more dynamic and active than the document networks. Scholars communicate not only through formal publications and conference proceedings, but also through email discussions, informal meetings, and chatting over lunch and coffee breaks. These channels of informal communication compose a rich array of information flows, which

foster the creation of original ideas and help mature scientific theories. The intrinsic limitations on formal and semi-formal document networks call for more advanced technology to materialize some of this less formal communication and exploit its richness, and more advanced social theory to understand its structure. - Figure 5 shows a social network formed by including informal communication channels. The connections are bi-directional, the in- and out- degrees follow a power-law distribution, and the network has a high cluster and low diameter, possessing the "small-world" attributes [29, 52].

We are beginning technical work that will permit the materialization and use of a hybrid network composed of both various formats of documents and social actors (authors and researchers). Different from the dual-layer network model advanced by Leydesdorff [35], this hybrid network links documents and social actors together. The hybrid network conceptualization facilitates the interaction between documents, social actors, and the combination of the two through various network tools. It takes advantage of both the social network and document network, thus providing the

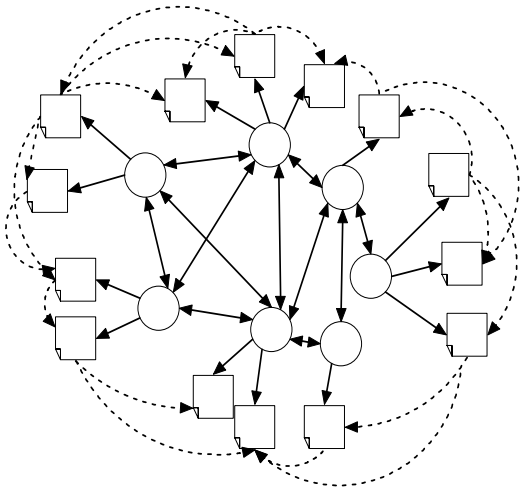

Figure 6. Hybrid Network

opportunities to design better tools to facilitate search and selection of both information artifacts and authors/researchers [24] (Figure 6). Our research will reveal the detailed structure of this hybrid network, methods for analyzing it, and mechanisms for using it to assist users of information systems.

Building on the materialization of this hybrid network, we plan a number of corollary investigations. We will capture the evolution of citation and hyperlink patterns among those formal, semi-formal, and informal communications. We will then explore the correlation between the citedness of a document and its quality and impact, and the different network measurements (centrality and prestige) of a researcher and his/her reputation and popularity. Furthermore, we will explore the differences between formal citation in official scholarly publication venues and hyperlinks in informal documents by extending evaluative bibliometics to semi-formal and informal scholarly communication. Usage analysis among these different types of documents will provide us with more insights into the emergence of authority and quality [10]. Kim [28] shows the motivation for linking is also multi-dimensional. We hope to understand under which conditions citations and hyperlinks can be used as a measure of impact, quality, relevance, reputation, and possibly generate other types of hyperlink and citation measurements of these scholarly achievements.

3.2 Building the technical foundation for hybrid network research

As described in the previous section we are investigating the structure and utility of hybrid networks to improve our understanding of scholarly activity, and as a tool that scholars can use to enhance their intellectual activity. In this context, we are investigating and implementing a number of technical extensions to scholarly communication systems. These efforts will require innovations in reputation management, recommendation systems, hybrid network visualization, and distributed document management. This technical work falls into two broad classes:

- *Expanding the breadth of documents* - Our goal is to enrich the structure of the hybrid document/social network with more types of documents. In addition to traditional formal publications, we intend to include annotations, reviews, recommendations, similarity-maps, data visualizations, and other forms of more personal communication that are often at the heart of scholarly activity. Our work builds on the foundation provided by arXiv and OAI-PMH.
- *Expanding the depth of documents* - Our goal is to enrich the structure of the hybrid document/social network by incorporating more complex and active *virtual documents*. These virtual documents exploit a number of key features including 1) a*ggregation* - they are composed from several sources of data of mixed genre; text, images, video, audio, database access, and other types; 2) *distribution* – the data in a virtual document may be locally "contained" in the document or contained by reference to external data sources; and 3) e*xecution* – the presentation of a virtual document occurs on the fly via the integration and execution of local or distributed parameterized services (programs) that process the local or distributed data that is "contained" in the

virtual document, producing on-demand *disseminations*. Our own work with Fedora (Flexible Extensible Digital Object Repository Architecture) [43, 44] in the context of digital library research provides the strong basis for our proposed investigations in this area. We are planning to integrate this digital library-based work into the context of the scientific research community, where Grid middleware [18, 19, 51] provides the basis for coordinating distributed computational resources and data.

4 Conclusion

This paper has described work on future scholarly publishing technology and models that we have initiated and are planning at Cornell Information Science. In the spirit of the broader information science program at Cornell that "studies information in its human and social context" [1], our work integrates both technology that underlies scholarly communication and the social organizations that this technology enables and encourages.

Acknowledgments

Although this paper has a single listed author, the ideas originate from a team of collaborators including Shay David, John Erickson, Geri Gay, Johannes Gehrke, Bing Pan, Sandy Payette, Jayavel Shanmugasundaram, Herbert Van de Sompel, and Simeon Warner. Support for the work described here comes from a variety of sources including National Science Foundation Grant No. IIS-9817416, National Science Foundation under grant number 0127308, and the Andrew W. Mellon Foundation.

References

[1] *Cornell University: Informatioin Science Program*, 2004 http://www.cis.cornell.edu/infoscience/.

[2] *Dublin Core/MARC/GILS Crosswalk*, 1997 http://lcweb.loc.gov/marc/dccross.html.

[3] *The Fedora™ Project: An Open-Source Digital Repository Management System*, 2003 http://www.fedora.info.

[4] *Gnutelliums*, 2002 http://gnutella.wego.com/.

[5] *OAIster...find the pearls*, 2002 http://oaister.umdl.umich.edu/.

[6] *Semantic Web Activity: Resourse Description Framework (RDF)*, 2001 http://www.w3.org/RDF.

[7] B. Ahlborn, W. Nejdl, and W. Siberski, "OAI-P2P: A Peer-to-Peer Network for Open Archives," presented at Workshop on Distributed Computing Architectures for Digital Libraries ICPP, Vancouver, B.C., 2002.

[8] R. Bachmann-Gmuer and P. Mueller, *mies - Meta Information Exchange System*, 2003 http://wymiwyg.org/mies/index.html.

[9] C. W. Bailey, Jr., *Scholarly Electronic Publishing Bibliography*, 2003 http://info.lib.uh.edu/sepb/sepb.html.

[10] C. L. Borgman and J. Furner, "Scholarly communication and bibliometrics," *Annual Review of Information Science and Technology*, 36, pp. 3-72, 2002.

[11] A. Buckholtz, "Returning Scientific Publishing to Scientists," *Journal of Electronic Publishing*, 7 (1), 2001.

[12] M. K. Buckland, "What is a "digital document"?," *Document Numerique*, 2 (2), pp. 221-230, 1998.

[13] C. Buckley, M. Burright, R. S.-W. Prendergast, and A. Taylor, "Electronic Publishing of Scholarly Journals: A Bibliographic Essay of Current Issues," *Issues in Science and Technology Librarianship*, Spring, 1999.

[14] I. Clarke, T. Hong, W., S. G. Miller, O. Sandberg, and B. Wiley, "Protecting Free Expression Online with Freenet," *IEEE Internet Computing*, 6 (1), 2002.

[15] A. Crainiceanu, P. Linga, J. Gehrke, and J. Shanmugasundaram, "Querying Peer-to-Peer Networks Using P-trees," presented at VLDB, Berlin, 2003.

[16] B. Cronin, *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham, 1984.

[17] N. Daswani, H. Garcia-Molina, and B. Yang, "Open Problems in Data-Sharing Peer-to-Peer Systems," presented at International Conference on Data Engineering, San Jose, 2002.

[18] I. Foster, C. Kesselman, J. Nick, and S. Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration," Global Grid Forum June 22 2002.

[19] I. Foster, C. Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," *International Journal of Supercomputer Applications*, 15 (3), 2001.

[20] E. Garfield, *Citation indexing: Its theory and application in science, technology, and humanities*. New York, NY: John Wiley, 1979.

[21] E. Garfield and A. Welljams-Dorof, "Citation data: Their use as quantitative indicators for science and technology evaluation and policy-making," *Science*

and Public Policy, 19 (5), pp. 321-327, 1992.

[22] E. Garfield and A. Welljams-Dorof, "Of Nobel class: A citation perspective on high impact research authors," *Theoretical Medicine*, 13 (2), pp. 117-135, 1992.

[23] J. W. Houghton, "Crisis and Transition: The Economics of Scholarly Communication," *Learned Publishing*, 14 (July), pp. 167-176, 2001.

[24] A. Hyatt, N. Contractor, and P. M. Jones, "Computational organizational network modeling: Strategies and an example," *Computational and Mathematical Organizational Theory*, 4, pp. 285-300, 1997.

[25] A. Johns, *The nature of the book : print and knowledge in the making*. Chicago, Ill.: University of Chicago Press, 1998.

[26] R. K. Johnson, "Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication," *D-Lib Magazine*, 8 (11), 2002.

[27] J. Kahan, M.-R. Koivunen, E. Prud'Hommeaux, and R. R. Swick, "Annotea: An Open RDF Infrastructure for Shared Web Annotations," presented at WWW10, Hong Kon, 2001.

[28] H. J. Kim, "Motivations for hyperlinking in scholarly electronic articles: A qualitative study," *Journal of the American Society for Information Science*, 51 (10), pp. 887-899, 2000.

[29] J. Kleinberg, "Navigation in a small world," *Nature*, 406, pp. 845, 2000.

[30] R. Kling and G. McKim, "Scholarly communication and the continuum of electronic publishing," *Journal of the American Society for Information Science*, 50 (10), pp. 890-906, 1999.

[31] R. Kling, G. McKim, and A. King, "A Bit More of It: Scholarly Communication Forums as Socio-Technical Interaction Networks," *Journal of the American Society of Information Science and Technology*, 54 (1), pp. 47-67, 2003.

[32] C. Lagoze, W. Arms, S. Gan, D. Hillmann, C. Ingram, D. Krafft, R. Marisa, J. Phipps, J. Saylor, C. Terrizzi, W. Hoehn, D. Millman, J. Allan, S. Guzman-Lara, and T. Kalt, "Core Services in the Architecture of the National Digital Library for Science Education (NSDL)," presented at Joint Conference on Digital Libraries, Portland, Oregon, 2002.

[33] C. Lagoze and H. Van de Sompel, *The Open Archives Initiative Protocol for Metadata Harvesting - Version 2.0*, 2002 http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html.

[34] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, *Implementation Guildelines for the Open Archvies Initiative for Metadata Harvesting*, 2002 http://www.openarchives.org/OAI/2.0/guidelines.htm.

[35] L. Leydesdorff, "Theories of citation," *Scientometrics*, 43 (1), pp. 5-25, 1998.

[36] C. Lynch, "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis on the Central Role of Trust," presented at Authenticity in a Digital Environment, Washington, D.C., 2000.

[37] P. Maniatis, M. Roussopoulos, T. J. Giulli, D. S. H. Rosenthal, M. Baker, and Y. Muliadi, "Preserving Peer Replicas by Rate-Limited Sampled Voting," presented at Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, Bolton Landing, NY, 2003.

[38] N. Medeiros, "E-Prints, Institutional Archives, and Metadata: Disseminating Scholarly Literature to the Masses," *OCLC Systems & Services*, 19 (2), pp. 51-53, 2003.

[39] W. Nejdl, B. Wolf, and C. Qu, "EDUTELLA: A P2P Networking Infrastructure Based on RDF," presented at WWW2002, Honolulu, 2002.

[40] J. J. O'Donnell, *Avatars of the word : from papyrus to cyberspace*. Cambridge, Mass.: Harvard University Press, 1998.

[41] C. Oppenheim, "The Future of Scholarly Journal Publishing," *Journal of Documentation*, 56 (4), pp. 361-398, 2000.

[42] A. Oram, *Peer-to-peer : harnessing the benefits of a disruptive technology*, 1st ed. Beijing ; Cambridge Mass.: O'Reilly, 2001.

[43] S. Payette and C. Lagoze, "Flexible and Extensible Digital Object and Repository Architecture," presented at In Research and Advanced Technology for Digital Libraries, Second European Conference, ECDL '98, Berlin, Heidelberg, 1998.

[44] S. Payette and T. Staples, "The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services," presented at European Conference on Research and Advanced Technology for Digital Libraries, Rome, 2002.

[45] S. Ratnasamy, S. Shenker, and I. Stoica, "Routing Algorithms for DHTs: Some Open Questions," presented at IPTPS, 2002.

[46] V. Reich and D. S. H. Rosenthal, "LOCKSS: A Permanent Web Publishing and Access System," *DLib*, 7, 2001.

[47] A. Rowstron and P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer

systems," presented at 18th IFIP/ACM International Conference on Distributed Systems Platforms, 2001.

[48] T. Staples and R. Wayland, "Virginia Dons FEDORA: A Prototype for a Digital Object Repository," *D-Lib Magazine*, July, 2000.

[49] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," presented at ACM SIGCOMM, 2001.

[50] A. Strauss, "A Social World Perspective," *Studies in Symbolic Interaction*, I (1), pp. 119-128, 1978.

[51] S. Tuecke, K. Czajkowski, I. Foster, J. Frey, S. Graham, C. Kesselman, T. Maquire, T. Sandholm, D. Snelling, and Vanderbilt, "Open Grid Services Infrastructure (OGSI): Version 1.0," Global Grid Forum draft-ggf-ogsi-gridservice-33, January 27 2003. http://www.ggf.org/ogsi-wg.

[52] D. J. Watts, *Six Degrees: The science of a connected age*. New York: W.W. Norton & Company, 2003.

[53] R. Williams, R. Moore, and R. Hanisch, "A Virtual Observatory Vision based on Publishing and Virtual Data," US National Virtual Observatory 2003. http://bill.cacr.caltech.edu/usvo-pubs/files/VO-vision.pdf.

[54] L. L. Zia, "The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program," *D-Lib Magazine*, 8 (11), 2002.