

SSRC / DCIP Digital Cultural Institutions Project

**Project APPRAISE: A framework for
understanding the hybrid structure of
online scholarly networks and reputation
economies**

Interim project summary and think-piece

By Shay David,
Science and Technology Studies Department /
Information Science Program
Cornell University

301 College Ave.
Ithaca, NY 14850

mailto: sd256@cornell.edu

Introduction - The world of scholarly publishing is changing and so are the challenges it faces

For nearly 350 years, the printed journal has been the core of a scholarly network aimed at communicating the results of scholarly activity, registering new results, resolving questions of originality and precedence, establishing the quality of results and ideas through peer-review, and awarding promotion and tenure. This system is now under serious internal pressure and is close to, if not beyond, the point of meltdown: its economic viability is questioned as rising subscription prices challenge flat library budgets; scholars are increasingly resistant to surrendering copyrights to publishers; and latency between submission and publication is at odds with the increasingly fast pace of scholarly innovation. Moreover, adding to the heat from the outside are the opportunities afforded by the Web and new digital publishing and information technologies that provide open and democratic alternatives to the traditional, closed, controlled, for-profit publishing model. In sum, with this implosion near, it is evident that the centuries old world of scholarly publishing is about to dramatically change. In fact, it has already dramatically changed. However, like all revolutions this ‘revolution’ too comes with a price tag. While novel technologies for the packaging and presentation of scholarly results alleviate some of the concerns that bring about the collapse of the traditional scholarly communication networks, equally daunting challenges are introduced. As the original system of scholarly communication is replaced by networked, digital, open, democratic infrastructure it is essential that we understand the functions that the old system fulfilled beyond the mere dissemination of information. It is of prime importance that we make sure that there are supporting mechanisms to allow at least some of those functions to adopt and to survive, lest the new systems be built on the same unstable ground. This short paper aims to highlight some of these challenges and to suggest an analytical framework that would allow us to rethink the basic concepts of authority, expertise and reputation as new digital cultural institutions are coming into being.

Roosendaal and Geurts used the metaphor of scholarly activity as *agora*, equating scholarly communities to an assembly for the generation of ideas. (Roosendaal and Geurts 1997) Clearly, the nature of these communities changes as the technology and nature of their documentation results changes. The exponential growth of information and the low cost of digital modes of communication create endless opportunities for the creation and publication of new ideas—in principle, ‘anyone can be an author’—and yet just like in the analog world only certain documents become the

center of their respective communities and only certain authors gain prominence. What, then, is the relationship between communities and documents in this digital and open environment? What are the social and cultural mechanisms that control inclusion or exclusion from a scholarly community and that govern reputation in an era in which technology allows almost anyone to ‘publish’?

These questions are not new. For more than a decade, significant attention has been given to all aspects of the development of digital, electronic, and online publishing alternatives to conservative press. A large body of literature aims to understand the implications of new modes of information exchange and their impact on the ways in which ideas are communicated and substantiated. Some authors have considered digital text and electronic publishing as an alternative to conventional print in a historical context (Harnad 1991; Bolter 1991). Others investigated the applicability of new multimedia and electronic print to educational environments (Ryder. 1995; N.C. Burbules and T.A. Callister 1996; Duncan 1997). Another body of work surrounds the relationships and interplays between writers and readers in cyberspace (Romano 1996; Roberts 1997; Lankshear 1997). And yet other important works chose to look at the battles surrounding the nature of the book and its future as a cultural artifact in the e-book form. (Lynch 2001) In addition, many accounts have addressed the merits of digital libraries covering both social and technological aspects (Arms 2000; Bhargava 2000; Chen 2001; Greenstein 2002; Pace 2003; Witten and Bainbridge 2002)

Many of the earlier accounts looked at advancements in technology in revolutionary terms: according to such (simplistic) accounts digital and electronic publishing offer unprecedented opportunities for scholars when they search for new ideas, formulate their own, disseminate information to peers and advance their research (Harnad, for one, talks about ‘electronic skywriting’ as a revolution comparable to the invention of print (1990)). More recent work, in contrast, departs from such technological determinism and tries to look at these new technologies with a critical eye and advocates that they be understood “not merely as technical developments, but as social, political, cultural, and economic phenomena” (Roberts 1999) Almost needless to say, that I espouse the latter approach and, in the tradition of science studies, focus on actual controversies as a way to gain insights into the practicalities of the phenomena. In my research I aim to unpack some of the complexities by looking at concrete case studies; as proposed below, a new approach that looks at scholarly communities as hybrid networks can shed some light on these matters. Moreover, with my peers at Cornell’s Information Science Program in general, and the APPRAISE project team specifically (see appendix below), I aim to develop a process of critical socio-

technical design practice that integrates technology production with a critical awareness of how technology shapes and is shaped by our culture. The results of this work, so I hope, will bring about not only better understandings of how the world of scholarly publishing is changing, but also expose the values embedded in the brave new world of online communication, and offer prototypes for systems that will be truly democratic and open.

Openness and reputation

To better understand the issues at stake we can scrutinize the institutions that have developed in recent years and that challenge traditional scholarly journals and paper-publishing. Among others, these alternatives include electronic repositories that facilitate democratic, open, and rapid publication and access to new results. Two quintessential examples for such institutions and the set of problems they bring about are PLoS - the Public Library of Science (<http://www.plos.org>) and arXiv (<http://www.arxiv.org>), the key site for my research. Encouraged by the earlier success of PubMed Central, and arXiv PLoS was founded in 2000, with the venerable aim of making the world's scientific and medical literature a public resource. PLoS journals are slowly becoming top-tier venues for publication but in order to gain momentum and reputation they needed to attract authors that have made names for themselves in the traditional print venues. For this reason the first issues of PLoS enlisted authors that have had previous publications in prestigious journals like Science and Nature suggesting that the 'revolution' might not be as swift as intended and that departing from earlier models of 'expertise' might take more time than initially estimated. Likewise, as discussed in further details below, arXiv is now facing challenges in the definition of reputation and expertise and has recently implemented quality control mechanisms that seem at odds with its open and democratic aspirations.

arXiv was historically designed as a means of automating an existing and burgeoning culture of 'pre-prints' in physics. Pre-prints are pre-refereed papers which report on new research, record precedence (who discovered what first) and solicit comments, preparing a paper for 'formal' publication. Pre-prints were initially distributed in paper format in a process that was much faster than traditional publishing but was still too slow and cumbersome. An e-print archive was designed to upscale and accelerate the process. While other disciplines like business and economics have their parallels of pre-prints, arXiv which was started at Los-Alamos by renowned physicist Paul Ginsparg, with High Energy Physics and evolved to include other branches of physics, mathematics, and computer science is one of the only high impact systems of its kind. In fact arXiv is today the most widely-used pre-print repository in all the fields of

science and scholarship. In its 13 years of operations it amassed over 289,000 papers (September 2004), and its repository is growing by several thousand publications per month.

Until recently, any author with a university affiliation could have published on arXiv. Naturally, arXiv's unorthodox mechanism of submission and peer review brought up a set of questions concerning the way by which the system should filter content. Scholars wondered whether "any attempt [should] be made to monitor quality? If quality *is* monitored, what criteria are to be used? Will users see this as unnecessary 'control'?" (Pinfield 2001) Recently these questions transgressed academic circles and exploded when Robert Gentry, a creationist, sued the operators of arXiv claiming that they refused a series of ten of his papers because of their religious content. In their defense arXiv's operators claimed that the papers were not scientific. (Brumfiel 2002) In essence the court was asked to intervene and decide between the competing values that underwrite arXiv—a democratic and open scholarly publishing environment that wants to remain open to even the most radical and innovative ideas, but which, at one and the same time, wants to bar the door on non-scientific research.

In two other recent cases, arXiv came under attack from two more unexpected directions when published content was found to be libelous in one case and plagiarized in another, and had to be removed. (Giles 2003) As a consequence, departing from its earlier laissez faire content management policy, the operators of arXiv introduced a new system of 'endorsement' that unlike that of double-blind peer review is based on personal acquaintance and aims to ensure that only content from recognized authors and/or from reputable institutions is published. The primary operating principle of this system is that new authors need to be 'endorsed' by recognized scholars that gained their endorsement-granting-status through past contribution to the relevant scientific field or through affiliation with a recognized institution. Prior to publication would-be authors are now required to submit a system assigned alpha-numeric endorsement code to a recognized pundit who can accept or reject the endorsement request. Younger scholars are thus explicitly encouraged to exploit their social connections as part of their indoctrination into a field.

Clearly the new endorsement system and the court cases signify tensions that lead to some very important questions. Is it the case, as has been suggested in the past (i.e. (B.A. Nardi, S. Whittaker, and Schwartz 2000)) that even in digital scholarship what is important is not what you know, but who you know? How can an open system maintain quality control and remain open? How are authorship, reputation, trust, and objectivity negotiated in such an open but regulated digital environment? How does the transition

from traditional publication institutions (taken broadly) to new ones affect the agency of the actors involved? And most importantly to our case here, what are the analytical tools that are at our disposal when we try to understand this system which is in radical flux?

A hybrid network approach

To start and answer some of these questions, we should look at documents, the artifacts of scholarly communication, and the ways they provide the context for ideation and other scholarly activities. I want to advocate the view that documents, taken broadly, are inseparable from the social networks in which they are embedded. This intricate relationship becomes a fundamental element in the creation of a reputation economy which is at the heart of scholarly activity, whether online or off. The interplay between information artifacts, the formations of scholarly communities, and their institutionalization, can be explicated in terms of a conceptual framework that is based on information flows. Within this framework we see all scholarly publishing and communication as information exchanges 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, minutes from informal meetings, or digital endorsements. Indeed, three classes of documents can be identified: formal documents, in the form of officially published journal papers and conference proceedings; semi-formal documents, such as the e-Prints in arXiv; and informal documents, such as email discussions, blogs, or even phone conversations and coffee-break chatter.

The official and semi-formal publication systems form document networks. Coexisting with the document networks, there are social networks composed of authors and researchers. Citation could traditionally only point backward in time, (Figure 1) however, the development of electronic publishing exemplified by arXiv has collapsed the time dimension of this citation graph: one hundred years ago you might only have been able to cite papers published in the past year; now you can cite contemporaneous papers. Substantial research has been done regarding why people cite each other. Leydesdorff (Leydesdorff 1998) discussed a dual layered network, including the social network of authors and network of their reflexive communications. 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, through the creation of informal affiliations and endorsement of reputation, 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. It is important to recognize how all three classes of documents play part in developing economies of reputation that work on several parallel levels. While the time dimension might have collapsed in arXiv, what remains intact is the need to establish common means of reputation exchange. As Masum and Zhang made plain in their *Manifesto for The Reputation Society*, such reputation exchanges have far reaching implications that extend well beyond scholarly publishing. (Masum and Zhang 2004). Surprisingly, however, unlike in the case of commercial services like those offered by Amazon, eBay, or Shopping.com (formerly ePinions), with a few exceptions in the scholarly communication networks little is offered in the way of tools that allow to manage reputation and expertise. One of the goals of the APPRAISE project (see below) is to change this situation by developing a set of open-source software tools that could be used both for social analysis and for the benefit of the users of venues like arXiv.

Conclusion

Combining social and document networks we can look at the scholarly community as a hybrid-network comprised of both documents and social actors (authors and researchers.) Extending existing social network analysis techniques with this new form of hybrid network analysis, which looks at both human and material actors (in a Latourian sense as expressed in (Latour 1987)) allows us to investigate the interactions of humans and digital artifacts in a useful way. Figure 2 depicts a social network that is formed by including informal communication channels. This hybrid network links together human and material actors and facilitates the interaction between documents and humans. An actor operating within this network can take advantage of both the social network and document network, and thus the network can facilitate improved search, discovery, selection, and reputation management of both information artifacts and authors/researchers (Hyatt, Contractor, and Jones 1997) (Figure 3). Adopting a hybrid-network analysis approach, I suggest, provides an opportunity to improve the design of mechanisms and tools for the newly formed digital cultural institutions to the extent that it makes both designers and users aware of the richness and importance of such hybrid interactions. Taken seriously, this approach can help us better understand the tensions between openness, regulation and reputation while starting to develop systems that address these challenges.

Appendix: APPRAISE – the larger project

APPRAISE is a three year project designed to address key aspects of this system in transition. As we further our understanding of the interaction between ideation, community and documentation, APPRAISE aims to gain social scientific insights and to develop new software tools as a means of promoting more dynamic interaction of documents and researchers. Using standard quantitative and qualitative research methods (including ethnography, social-network analyses, citation analysis) combined with hands-on open-source design and coding, the aim of APPRAISE is to reveal the detailed structure of this hybrid network while developing new methods for analyzing it and tools to improve it. APPRAISE involves people and resources from Cornell's Science & Technology Studies and Communication departments and from Cornell's Information Science program as well as people from various other institutions (like the National Virtual Observatory).

The intellectual merit and broader impacts of the proposed research derive from the contexts in which the research will take place; pursuing this research in collaboration with and with full cooperation of the arXiv personnel ensure applicability of APPRAISE results to real-world scenarios beyond the usual short funding cycles of similar research. I am indebted to the other APPRAISE team members Geri Gay, Carl Lagoze, Simon Warner, and Bing Pan for their ongoing contribution to this project.

Figures

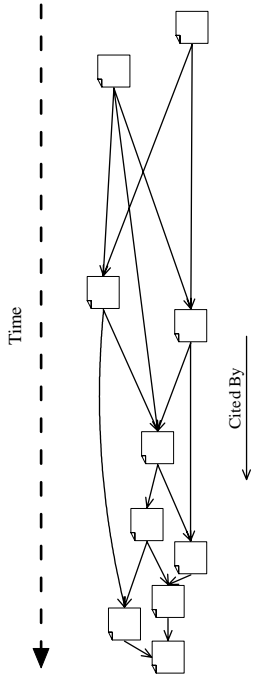


Figure 1. Scholarly Publishing Network

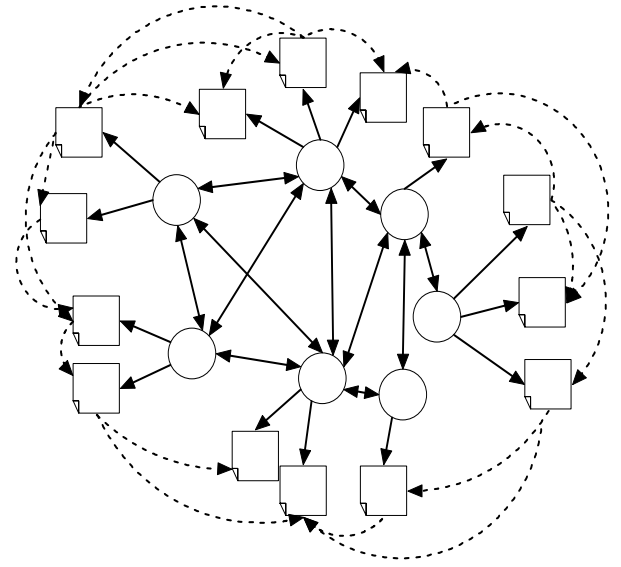


Figure 3. Hybrid Network

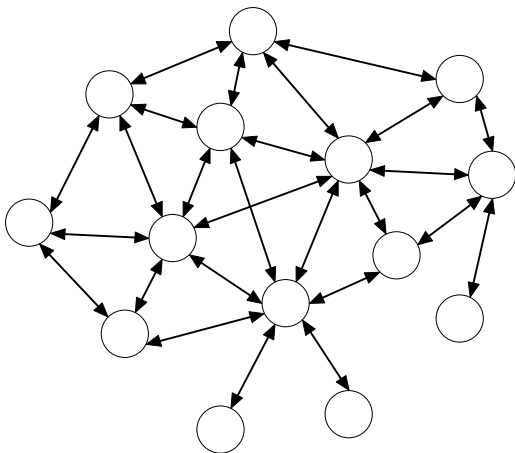


Figure 2. Researcher Social Network

References

- Arms, William Y. 2000. *Digital libraries*. Cambridge, Mass.: MIT Press.
- B.A. Nardi, S. Whittaker, and H. Schwartz. 2000. It's not what you know, it's who you know: Work in the information age. *First Monday* 5 (5).
- Bhargava, Bharat K. 2000. *Digital libraries and multimedia*. Boston: Kluwer Academic.
- Bolter, J.D. 1991. *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, N.J: Lawrence Erlbaum.
- Brumfiel, Geoff. 2002. Ousted creationist sues over website. *Nature* 420 (597).
- Chen, Ching-Chih. 2001. *Global digital library development in the new millennium : fertile ground for distributed cross-disciplinary collaboration*. Beijing, China: Tsinghua University Press.
- Duncan, B. 1997. Hypertext and education: (Post?)Structural transformations. In *Philosophy of Education Society Yearbook*, edited by S. Laird. Champaign, IL: PES Publications.
- Giles, Jim. 2003. Preprint server seeks way to halt plagiarists. *Nature* 426 (7).
- Greenstein, Daniel I. 2002. *The digital library : a biography*. Washington: Digital Library Federation and Council on Library and Information Resources.
- Harnad, S. 1990. Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry. *Psychological Science* 1 (6):342-344.
- . 1991. Post-Gutenberg galaxy: the fourth revolution in the means of production of knowledge. *Public-Access Computer Systems Review* 2 (1):39-53.
- Hyatt, A., N. Contractor, and P. M. Jones. 1997. Computational organizational network modeling: Strategies and an example. *Computational and Mathematical Organizational Theory* (4):285-300.
- Lankshear, C. 1997. *Changing Literacies*. Buckingham, UK: Open University Press.
- Latour, Bruno. 1987. *Science in action : how to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press.
- Leydesdorff, L. 1998. Theories of citation. *Scientometrics* 43 (1):5-25.
- Lynch, Clifford. 2001. The Battle to Define the Future of the Book in the Digital World. *First Monday* 6 (6).
- Masum, Hassan, and Yi-Cheng Zhang. 2004. Manifesto for the Reputation Society. *First Monday* 9 (7).
- N.C. Burbules, and Jr. T.A. Callister. 1996. Knowledge at the crossroads: Some alternative futures of hypertext learning environments. *Educational Theory* 46 (1):23-50.
- Pace, Andrew K. 2003. *The ultimate digital library : where the new information players meet*. Chicago: American Library Association.
- Pinfield, Stephen. 2001. How Do Physicists Use an E-Print Archive? Implications for Institutional E-Print Services. *D-Lib Magazine* 7 (12).
- Roberts, P. 1997. Literacies in cyberspace. *SET: Research Information for Teachers* (3).
- Roberts, Peter. 1999. Scholarly Publishing Peer Review and the Internet. *First Monday* 4 (4):2.
- Romano, S. 1996. Writerly readers: Close readings of close relationships. *Computers, Writing, Rhetoric and Literature* 2 (1).

- Roosendaal, Hans E., and Peter A. Th. M. Geurts. 1997. Forces and functions in scientific communication: an analysis of their interplay. Paper read at CRISP 97 Cooperative Research Information Systems in Physics.
- Ryder., M. 1995. Production and consumption of meaning: The Interplay between subject and object in open hypertext representation. Paper read at Semiotics as a Bridge Between the Humanities and the Sciences, 2-5 November 1995, at Victoria College, University of Toronto.
- Witten, I. H., and David I. Bainbridge. 2002. *How to build a digital library*. San Francisco, Calif.: Morgan Kaufmann.