

nature

webdebates

## Evolution and scientific literature: towards a decentralized adaptive web

Richard Luce

Director, Research Library of Los Alamos National Laboratory

The current debate over electronic access to primary literature is but the first wave in a sea of change transforming scholarly communication. Focusing solely on free versus fee makes for lively debate but it must be viewed in the context of this wider evolution. More is at stake than whether archives can be viewed 'for free' by the scientific community and society at large. From this viewpoint, the proposal by the [Public Library of Science](#), that one massive, centralized, 'free' repository will sufficiently meet end-user needs over time, appears as quite a gamble.

There are many factors driving the current crisis in scholarly communication. The current 'value chain' from author to reader for the creation and distribution of formal publications is long, and the many layers between input and output result in high costs and slow turnaround times. In addition to the editorial tasks involved in submission, editing, peer review and primary publication, the supplier chain includes interactions between primary publishers, abstracting and indexing services, secondary database publishers, distributors and libraries. Markets dictate efficiency and inefficiencies are usually corrected over time. It is reasonable to conclude that today's value chain is not sustainable in its current form; publishers, secondary providers, aggregators and libraries are already under siege, and market forces will drive substantial change in the roles they currently play.



Is the PLS proposal the solution? The free-enterprise model requires competition to allow the best solutions to emerge by stimulating the experimentation and urgency required to bring new ideas rapidly in to play. Experience has shown that centralized solutions may not be the best option for ensuring innovation over time. The notion of a centralized archive has some attractive advantages, notably the convenience of a one-stop shop and presumably serving as the enabling mechanism to access standardized metadata. But it also carries significant disadvantages. Centralized approaches are traditionally more vulnerable to failure, for example, by putting all one's eggs in the same basket, and a tendency to pick winners prematurely.

The shadow of politics looms over any initiative dominated by a single country and centralized repository approaches are typically discipline centric, which engenders problems categorizing new, trans-disciplinary science. PLS advocates cite the following advantages: the efficiency of large-scale searching on a single site; extensive citation interlinking between reports originally published in diverse journals; and linking to other types of data. Although such capabilities are clearly advantageous, they do not require a centralized approach and are available today, although certainly not 'free'. It is an opportune time to experiment and rethink the assumptions that underlie our systems. If we believe it is prudent to hedge our bets, many alternatives should be considered.

### A decentralized approach: the Open Archives Initiative

The revolution brewing in the formal system is centred on the goal of returning choice to authors. Efforts to give authors control over the communication and distribution of their work, in the form of author self-archiving systems, are gaining ground. Self-archiving allows authors to deposit their papers or preliminary drafts into a repository and thereby speed up the communication process, with submittal for publication and peer review following later at the discretion of the author. The [Open Archives Initiative](#) (OAI) was organized to create a technical forum to solve interoperability issues between various author self-archiving solutions. The initiative seeks to develop a framework for a distributed 'universal e-print archive' by establishing interoperability standards supporting the search and retrieval of e-print papers from

all disciplines. Protocols have been developed to ensure these archives work together so that any paper in any of these archives can be found from anyone's desktop, as if it were all in one virtual public library.

The objective of the OAI is to develop a framework to facilitate the discovery of content stored in distributed archives and to achieve interoperability among archives that cut across physical, organizational and disciplinary boundaries. The OAI framework supports both data providers (archives) and service providers, which develop value-added services based on the information collected from cooperating archives. The OAI is becoming widely accepted among a broad array of resources, with [26 registered conforming repositories](#) and many more expected in the near future after the experimental framework is tested.

Why is the OAI relevant to the PLS debate? It enables a wide choice of approaches for e-print archives to be created wherever they make sense, for example, by discipline, institution, society, country and so on, as well as allowing the free-market system to work among competing service providers. It provides service providers with the mechanism to collect openly available content, which can support cross-archive searching while inducing a free-market approach to innovation and the development of new services. Finally, it can coexist with the current formal system, allowing us time to evolve and mature our evolutionary direction and expertise. With the growing use of e-print archives, we are witnessing a transition from the old model of formal communication to a rapidly evolving hybrid.

## **An evolutionary direction - the adaptive web**

As evidence of the early stage of this evolution in scientific communication, it is noteworthy that the current systems and counter-proposals essentially duplicate the current print medium. Content, historically in the form of the published article, has been the primary unit of value. Increasingly, value resides not just in the papers but also in the relationships between the papers, the associated dialogue from comments and reviews, updates to the original work, and the ancillary supporting materials. Typically, when hypertext browsing is used to follow links manually for subject headings, thesauri, textual concepts and categories, the user can only traverse a small portion of a large knowledge space. To manage and utilize the potentially rich and complex nodes and connections in a large knowledge system such as the distributed web, system-aided reasoning methods would be useful to suggest relevant knowledge intelligently to the user.

As our systems grow more sophisticated, we will see applications that support not just links between authors and papers but relationships between users and information repositories, and users and communities. What is required is a mechanism to enable communication between these relationships that leads to information exchange, adaptation and recombination. A new generation of information-retrieval tools and applications are being designed that will support self-organizing knowledge on distributed networks driven by human interaction (see [Active Recommendation Project](#) at Los Alamos). For example, to support trans-disciplinary science we stimulate different databases to learn new terms and adapt existing keywords to the categories recognized by different communities. This capability would allow a physicist or chemist to collaborate with colleagues in the life sciences without having to learn an entirely new vocabulary.

Through the use of these new tools, we will derive a shared knowledge structure that is based on users and usage in addition to that provided by author citations. Thus, the aggregated connections that readers make between papers and concepts will provide an alternative conceptualization of a given knowledge space. Such techniques will be coupled with classical search and retrieval methods, and these capabilities have an obvious utility for discovering and supporting evolving knowledge from these networks.

This emerging adaptive web will analyse and use the collective behaviour of communities of users, utilizing concepts such as adaptive linking, which facilitates the evolution of knowledge structures based on collective user-behaviour over time, and spreading activation, which uses a memory-recall process model from cognitive psychology (1). For example, using known keywords to search across distributed open archives, a user would receive recommendations of other conceptually related keywords, relevant articles, data sets and so on, based on semantic proximities linked across a multitude of distributed information resources. At the same time, the knowledge system the user has interacted with can begin to reorganize itself by incorporating

feedback from the interaction into its knowledge structure.

From the user perspective, such systems can use adaptive webs as a communication fabric to manage and co-evolve the knowledge traded with communities of members and users. Correspondingly, these new tools and systems will influence the adaptation of the structure and semantics of scientific discourse. Many questions remain unresolved, such as how we evaluate the knowledge structures and representations of such size and complexity.

Perhaps it is time to apply a lesson from biological diversity and stimulate our complex scholarly communication system to adapt in a multitude of new ways to benefit the research community by seeding and funding open and distributed resources. The PLS centralized approach is unnecessarily constrained in this context. A key notion from complex systems theory is that of many simple processes, under selective pressures, being able to interact synergistically to produce desirable global behaviour. This notion can be applied successfully to different problems and the scholarly communication system is a prime candidate.

Recent notable examples where decentralized efforts have succeeded with innovative approaches include diverse experiences such as decoding the human genome, the open source movement and peer-to-peer networks. It would be in our best long-term interests to optimize our communication systems to support a variety of approaches while we evolve our understanding of the coming adaptive web and its impact on the communication of science.

#### References

1. Rocha, L. & Bollen, J. *Biologically motivated distributed designs for adaptive knowledge management*. In *Design Principles for the Immune System and other Distributed Autonomous Systems* (eds Cohen, I. & Segel, L.) (Oxford University Press, 2001).

