

Reflections Towards the Development of a “Post-DL” Research Agenda

Clifford Lynch
Draft of June 10, 2003

I am struck by how greatly the DL program has proceeded by analogy to traditional libraries and library use; indeed, as a consequence of this considerable effort has been expended in attempting to understand the relationships and connections between new digital libraries and libraries as we have traditionally understood them.

One can think of a spectrum from modernization through innovation to transformation which describes how new technologies can be employed to change human behavior, societies, and organizations. Along this spectrum, much of the work on digital libraries has emphasized modernization – applying technology to do what we have always done, only more efficiently and effectively.

This is not intended to belittle the very real and significant accomplishments of digital libraries or to suggest that there isn't still a tremendous amount of valid and important research still to be done on them. Well-designed and well-run digital libraries can make an enormous difference for their user communities. But we should recognize the limitations of a research program focusing on digital libraries as we understand them today. This is likely to lead to mostly incremental rather than transformative progress.

In order to define a new research program for the “post-DL” era, one that builds upon and integrates the existing work on digital libraries as well as the enormous secular changes that have taken place in human and societal behaviors and aspirations within the context of the pervasive deployment and continuing progress in information technology and networked information, one useful strategy may be to step back. Rather than considering how to re-design or recreate or enhance libraries as digital libraries we might usefully focus our attention on the human and social purposes and needs that libraries and allied cultural memory institutions have been intended to address – recognizing that they are not the exclusive agencies addressing such purposes and needs, and recognizing further that there are closely-related information management purposes and needs both long-standing and newly-emergent that have not been satisfactorily addressed by cultural memory organizations, or indeed, by any other players.

In reviewing this collection of information management purposes and needs, we face a serious scoping problem which I will largely ignore here, though I think we may well find it useful to explore it in our deliberations at the upcoming workshop. It would be easy to draw the boundaries too broadly, with a loss of focus, or too narrowly, with a loss of holistic perspective.

Libraries and related cultural memory organizations developed in response to individual and social needs to:

1. Ensure that information, knowledge, evidence, and discourse are collected, curated and preserved, for use today and in the future.
2. Ensure access to organized bodies of information, knowledge, evidence, and discourse.
3. More controversially, to provide some level of selection, filtering, and validation of these materials.

Further, libraries exist within a complex and continuously-evolving knowledge ecosystem that encompasses the lifecycle of information and knowledge from creation through dissemination and curation to use. It includes activities that we have historically called scholarly communication and the dissemination and use of cultural materials. Cultural memory organizations provide an essential foundation for this system but also respond to the ecosystem's characteristics more than shaping these characteristics. *It is understanding this ecosystem broadly and the potentials for its transformation which should be the focus of a research program, and not simply the historic or emerging roles of libraries within it.* And we must be careful not to overly-emphasize the parts of this knowledge ecosystem that are familiar, that we are comfortable with intellectually, socially and economically, to the exclusion of the new, the unfamiliar, the disturbing, the confusing. The research program must go beyond the naturally conservative inclinations of most cultural memory institutions to actively survey, explore and engage the new and most dynamic parts of the ecosystem.

Finally, we should recognize that there are now vast unmet needs; they have been largely ignored or defined as out of scope by the existing players in the knowledge ecosystem, and largely unaddressed either by research or by commercial product developments. These unmet needs have become acute due to the implications of the explosion of networked information, personal computing, massive and affordable computational and storage resources, emerging sensor and capture technologies, and from the increasing centrality of data, information, and knowledge in personal, organizational and social practices. The much discussed problem of "information overload" is one cryptic but popular marker for these developments.

These purposes and needs at both the individual and social levels help to identify what might be some of the key topics of a "post-DL" research agenda.

We need to incorporate a truly user-centric approach, recognizing that individuals now hold very large personal digital libraries and collections of records of interactions with other information services. It is only very close to the user that all of this comes together. Both individuals and the organizations and societies they participate in need help with the implications of the existence of such collections, and help in managing and exploiting them effectively. The paper by the DELOS/NSF working group on Personalization is highly relevant here, as is the thinking of the ARPA Life Log program (though I suspect that a post-DL research agenda would stress components of a Life Log related to information and information interactions, rather than panoptic sensing, and would be more focused on interactions among Life Logs, and between Life Logs and information resources.) We need to understand and explore the potential of the computer as assistant or collaborator in information-intensive activities and over long time horizons. Following on from this are a range of questions about how to structure and manage personal information over long periods of time (human lifetimes and beyond) and to integrate this usefully and respectfully with organizational and “public” information.

The processes of authoring and structuring information and knowledge and of information and knowledge use or re-use will be critical. We need to explore the implications of authoring or data structuring or knowledge representation practices for the conduct of science and scholarship, and for other fields of human endeavor. We need a much closer look at the processes and tools and social aspects of “authoring” and “reading” (in the broadest senses). We need to think about authorship practices in a world where less and less of the potential readers are human beings, and also about the reading practices and approaches of these non-human readers (for example, manipulation of structured information objects vs. computational linguistics approaches).

Social communities grow up around, interact with and structure information and knowledge; information comes from many sources, and is often contradictory, redundant or inconsistent. Tools to construct, analyze, model, simulate, and support social communities in conjunction with the information lifecycle are needed. We need to examine anew trust, reputation, belief, inconsistency and uncertainty in the distributed digital environment where assumptions about underpinnings such as identity are simultaneously being questioned. Network-based communities also interact with economics, business models and markets in ways that are not yet well understood.

Finally, there is the entire area of the stewardship, preservation and curation of information, discourse, knowledge, data and culture. There are tremendous technical, economic, legal and political problems here; much progress has been made in mapping these problems, but much less in developing solutions. And again, we need to translate these issues into a personal, user-centric perspective as well as exploring them within the existing institutional frames. We also need to consider these issues not only in the small but in the large – the potential

importance of stewardship, preservation and curation as public policy goals, and the relationships between these activities and national security, or the protection of a nation's cultural heritage, for example.

In the DL program, prototypes – including large-scale prototypes – have been very valuable. In a post-DL research program, however, in part because of the concern with very large-scale and long-time-horizon phenomena, the focus on prototypes will need to be complemented by a new investment on models and simulations, and also some accommodation of long-term research projects.

Why are these issues important? Why should we invest scarce research dollars here, and invite the engagement of the even scarcer resource in the form of our best researchers? These issues reach to the very heart of our ability to continue to function effectively and responsibly as individuals, as members of organizations, and as a society across decades and generations. They address fundamental unmet needs of individuals, organizations and societies; research breakthroughs in these areas can enhance our ability to conduct scholarship and science, improve education and learning, make our industries and government more effective and more competitive, and give birth to entire new technology-driven industries. They can also ensure that our cultural memory organizations can continue to evolve and function in a responsive and appropriate way.