

Advancing Scholarship & Intellectual Productivity

AN INTERVIEW WITH CLIFFORD A. LYNCH (PART 1)

By Brian L. Hawkins



Clifford A. Lynch, Executive Director of the Coalition for Networked Information (CNI), is the recipient of the 2005 EDUCAUSE Award for Leadership in Public Policy and Practice, sponsored by SunGard SCT, an EDUCAUSE Platinum Partner. A joint initiative of the Association of Research Libraries (ARL) and EDUCAUSE, CNI is dedicated to supporting the transformative promise of networked information technology for the advancement of scholarly communication and the enrichment of intellectual productivity. (For more information on CNI and its work, visit <http://www.cni.org>.) Lynch is a leading analyst of technology trends in education, libraries, publishing, and cultural arenas and an advisor to organizations and leaders involved in those fields. Before joining CNI in 1997, he spent eighteen years at the University of California Office of the President, the last ten as Director of Library Automation. For well over twenty years, he has been at the leading edge, developing studies, practices, and policies that have helped library and IT professionals understand and employ significant innovations in making networked information available to faculty, students, and the general public, with particular attention to the policy implications of technical issues.

After CNI's fall Task Force Meeting held in Phoenix in December 2005, **Brian L. Hawkins**, President of EDUCAUSE, sat down to talk with Lynch about—among other subjects—the current CNI initiatives, digital assets preservation, e-science, institutional repositories, Google, and future challenges for networked information technology.



Hawkins: Clifford, how did you get your start in this field, one that is only now starting to be defined?

Lynch: As your question suggests, it's hard to characterize how I got a start in a field when it wasn't a field when I started. I began my working career in the mid-1970s doing what was then called *library automation* or *library systems*. I did a variety of things, including a stint in academic computing at NYU, which eventually took me to the University of California system when it was building its union online catalogue back in 1979. That led into a broader engagement with networking. At the University of California, at least, library automation turned out to be a big driver for intercampus networking, even before the Internet became widespread.

Hawkins: When did you start working with CNI, and how did you become associated with CNI?

Lynch: The short answer is that I became the director of CNI in mid-1997, but there is also, of course, a much longer answer. CNI was founded in 1990. However, CNI has a prehistory that is not widely known. Starting in the mid-1980s or so, a group of key institutions—notably Penn State, UC, and Virginia Tech—came together, informally and periodically, under the leadership of a group of now-legendary figures that included Richard West, Bob Heterick, Gary Augustson, Nancy Cline, Paul Gherman, Marilyn Sharrow, and others. They did some joint networked information R&D projects, had some interesting interactions with DEC Research,

California, which is an astounding collection of talent—both the team at the Division of Library Automation, where I worked, and my colleagues throughout the campuses. Richard West was certainly an important influence, as was also Michael Buckland, now professor emeritus at Berkeley's School of Information (where I'm adjunct faculty). I first got to know Michael when he was the Assistant Vice President of Library Plans and Policy at UC in the 1980s. He has been incredibly important to me as someone I've learned a great deal from and as a collaborator I can test and explore ideas with. Michael has an amazing grasp of the deep history of libraries and cultural memory organizations and of how that history can inform us about the present and the future.



"At the University of California, at least, library automation turned out to be a big driver for intercampus networking, even before the Internet became widespread."

Until I left in 1997, at the University of California system, intercampus data networking was a library automation function. As Director of Library Automation, I held responsibility for intercampus data networking, which is not usually the way that function is handled, though at the time, library automation did report up to the CIO.

My academic training included an M.S. from Columbia University in computer science, and a lot of my computer science work—particularly my doctorate from UC Berkeley in the mid-1980s—involved information retrieval and data management. And then, in the 1990s, all these threads came together for me when what had been library automation, concerned mostly with systems like catalogs and indexes for managing and providing access to enormous collections of physical objects, morphed—under the influence of Moore's law, the ever-more-capable and ubiquitous Internet, and other factors—into a set of activities increasingly involved with *digital* content. In my mind, this is where the whole world of networked information truly launches from.

and mapped out some of the early agenda of CNI. When CNI became an official entity as a joint project of Educom, CAUSE, and the Association of Research Libraries, Paul Evan Peters was appointed as the founding director. I was involved heavily in CNI throughout those initial years, in project groups, task forces, and other activities. Paul died in 1996, and I came on as director in 1997. So, I do feel like I've been around since the early days or even before the early days of CNI.

Hawkins: Can you tell us the key people who have influenced you in your career—and why?

Lynch: That's a rather long list, and there is no way I can do justice to all the people I owe. I had many wonderful, inspiring teachers at Columbia and later at Berkeley. I was fortunate to work for a long time for Ed Brownrigg, who was a very important visionary and leader in library automation in the 1980s and early 1990s. I learned an incredible amount from Ed. I was influenced by many people at the University of

There are many others. My wife, Cecilia Preston, has been important in ways too numerous to recount here, not just as someone who has been there for me but as a major intellectual collaborator and influence. You yourself, Brian, have played an important role, as has Duane Webster at ARL. And of course, Paul Peters was very influential. He and I were colleagues and good friends, and we shared a huge collection of common interests, ranging from networked information all the way through science fiction and how it can help us to understand possible futures. I can't tell you how much I treasure the memory of some of my conversations with Paul about these questions.

I also want to be sure to acknowledge Joan Lippincott at CNI. So much of what CNI has accomplished over the years has been the result of her efforts, and since I've taken over as director there, virtually all the work I've done has been the result of an extensive collaboration and dialogue with her.

Hawkins: What do you see in the future for networked information technology?

Lynch: There's not one technology in question here; it's the convergence of many technological and social developments. There are many players. As I look at the future of many of these technologies, one of the big issues I see is the role of industry in their development—the extent to which industry specializes technologies to the scholarly world as balanced by the extent to which industry simply develops very broad commercial-market, off-the-shelf, consumer-market technologies. Figuring out how to engage all of the industry players, as that pendulum moves back and forth, is a significant issue. One of the things I struggle with is how much we are going to be able to afford to develop technologies specifically to meet the needs of the world of education and scholarship and how much we're going to have to rely on these general-purpose developments.

A case in point: Are we going to rely on Google to index everything, at a lowest-common-denominator level, and are we going to rely on Google's interest in developing projects, like Google Scholar, to address the needs of higher education, or are we in higher education going to supplement these developments with sophisticated finding tools in various scholarly disciplines? Certainly, there is a history in some scientific disciplines (for example, chemistry and molecular biology) of developing very sophisticated tools—tools that help scholars discover, navigate, and analyze data. This is going to be a big challenge in the future.

It is also striking to me that colleges and universities are doing a lot less in the way of local development. As a basis of comparison, in the 1990s, various institutions were conducting quite sizable software engineering projects; very few are doing so today. There has been a resurgence in higher education software development projects over the last couple of years, but their character is

quite different, illustrating how the costs and complexity of software development have ratcheted up. Today, higher ed software development involves big collective projects like Sakai, projects that represent investment collaborations among a whole constellation of colleges and universities, fueled further by grant funding. That's very different from the situation in the early 1990s, when there was a lot of localized but fairly extensive innovation going on, largely with local funding.

It's clear that in higher ed today, we need to pick our projects very carefully; we can't support many projects concurrently. The danger here is that this makes us much more conservative in picking our projects, because we can't afford to have any of them fail.

Hawkins: What are the key technology issues that need to be addressed?


Lynch: We need continued progress in the underlying technologies of storage and of data transmission. We need storage that is a lot faster, a lot cheaper, a lot more stable and robust. And we need still faster and still more ubiquitous networks. I'm increasingly struck by the gap between what a well-situated researcher at Stanford or Berkeley or MIT can do versus what someone can do from home. Broadband to the home is a real mess in this country; we have not done well on the public policy necessary to ensure we get what we need in this area—and there's so much at stake: the future of distance education, the ability to collaborate in a wide range of contexts, even the stewardship of personal collections of information that we need to be able to back up over these broadband connections, if only they were up to it. We also need to see some real breakthroughs in operating systems software—things that help us to organize and preserve and manage our computers in better ways and that make security more achievable for all users.

Technologies for knowledge representation and manipulation are still, in my view, in their infancy. Our ability to repurpose various kinds of data and information flexibly into different contexts and purposes is still very poor. Just think about the problems involved in harnessing and repurposing the array of existing digital scholarly resources that are relevant to doing something like a multiplayer simulation game of the political and economic alternatives facing the Roman Empire in the second century.

There are a couple of “Holy Grail” technologies that we don't yet have. We're getting pretty good at optical character recognition, at least for modern Romance texts—English, Spanish, French, Italian. We can even use OCR with older books that follow the typographic practices of the eighteenth century—though more research funding here would be very helpful. We do much less well in most of the other key languages, modern or ancient. In addition, we are getting surprisingly better at some parts of speech-to-text conversion. This starts to give us a way to organize and provide access to the spoken record. But our technological capability to understand video or still images is still in its infancy. Image interpretation is amazingly difficult and complex. Given that our cultural record is increasingly embodied in these images and sound recordings, making the technological breakthroughs to enable machines to characterize what's in images and videos for indexing, retrieval, organization, data mining, and similar purposes would have enormous implications.

Hawkins: Can you describe the current CNI initiatives?

Lynch: I can give you some highlights. Every year, CNI puts together a program plan, which is our vehicle for communicating our initiatives to representatives at member institutions and to other



“It's clear that in higher ed today, we need to pick our projects very carefully; we can't support many projects concurrently.”



“As we deal with scholarship underpinned by evidence that is extensively in digital form, the shape of the entire scholarly communication system is going to change.”

interested parties and which also, I hope, serves a secondary role by providing our member representatives with a way to communicate these initiatives more broadly within their institutions. One of the major themes we've sketched out in our 2005–2006 program plan is a focus on the management and stewardship of digital assets in institutional settings. This has an enormous set of ramifications and subactivities: support of electronic theses and dissertations; advancement of institutional repository services; records management in institutional settings; and outreach to organizations such as college/university museums to help them digitize their collections much as libraries have been digitizing their special collections.

Another area that is closely related and that has been very important to us over the years is digital preservation—or, to put it in another and perhaps clearer way, the preservation of material that enters life in digital form, as more and more of our scholarly and broader cultural materials do today. There are a tremendous number of difficult technical issues involved here. Even harder to address are the legal, economic, organizational, and other structural issues for how we, as a society, are going to deal with this flood of fragile digital content.

At CNI, we are also very interested in the ways in which changing scholarly practice—particularly the growing reliance on digital data and advanced information technology and collaboration tools—is affecting scholarly communication. Clearly, the traditional idea of communicating research in journal articles or, in the case of the humanities, in monographs is much too limited. As we deal with scholarship underpinned by evidence that is extensively in digital form, the shape of the entire scholarly communication system is going to change.

In addition, we've become very interested in the infrastructure supporting the research side of our colleges and universi-

ties. Certainly, there has been huge investment over the last few years in infrastructure to support teaching and learning: learning management systems; smart classrooms of various types; all kinds of new learning spaces and, particularly, informal learning spaces. CNI has spent much time on the question of what will happen to libraries and to campus public commons as we move into an increasingly digital world. In the late 1990s, a lot of institutions also invested heavily in an almost complete overhaul of administrative systems, partially driven by the Y2K threat. This proved to be a massive, and massively expensive, effort that I think took longer and cost more than most institutions expected—one that higher education is just now recovering from. The opportunity costs here were huge. Now, though, we're seeing a renaissance of interest in research computing. The economic, technological, operational, and policy trade-offs around centralization of computational cycles and storage have changed. Networking needs in research have also grown. Most important from CNI's point of view, however, is that research practices have changed quite a bit over the years. Research infrastructure is no longer about just cycles and networks. Today it is also about very-large-scale data management and about facilitating collaborations.

Other initiatives that are occupying our time at CNI include authentication and access management. These are central problems that continue to take on new dimensions. For example, we see a growing need to set up very lightweight collaborations across institutions, so-called *virtual organizations* that could be set up and torn down pretty much at the grassroots level to support collaborations. Access management technologies need to accommodate this.

Hawkins: What is the relationship among digital asset management, digital preservation, and the role of academic libraries?

Lynch: *Digital asset management* is an overarching term. I cringe a little bit whenever I hear the word *asset*, because it's accurate in some ways, but it also has a monetized context that feels a bit wrong when we start talking about the records of scholarly achievement and about the intellectual and cultural life within our higher education institutions. Sure, these are assets, but the word *assets* doesn't quite capture the importance of these things—the special place that they properly hold (or should hold, in my view) in our society. It doesn't recognize the stewardship *obligation*.

Increasingly, we're taking a lifecycle perspective on digital information objects and resources. Some assets have value for only a limited time and aren't really important to keep for the long run. *Digital preservation* is what we do with those digital assets that we indeed need to keep for protracted periods of time. Preservation deals with the long run. Note, however, that digital information has a unique profile of vulnerability and robustness that is very different from that of physical artifacts of various kinds. This profile includes crucial and too-often-overlooked issues of short-run management like backups and security. I don't think of these as being directly part of digital preservation, but they are part of the lifecycle management of digital materials, and they are certainly a necessary prerequisite to digital preservation. To put it another way, if the bits don't survive into next year, we won't have to worry about our ability to interpret them usefully two centuries from now.

Historically, at least in the United States, our great research libraries have born the brunt of the responsibility for managing our scientific and scholarly record. National libraries have played a role—the Library of Congress has certainly played an important role, especially in the humanities and some social sciences—but our true national libraries, such as the National Library of Medicine, have mandates that

cover only very limited parts of this landscape. The NLM and the National Library of Agriculture have done tremendously well within the scope of their mandates, but historically, the bulk of the burden for preserving the scientific record, in particular, has fallen to our research libraries, which are almost entirely embedded within universities. I think it's inevitable that these libraries will take the lead in preserving the record of science and scholarship as it moves to digital form.

Hawkins: What are the implications of e-science and cyberinfrastructure, particularly in the data and information management areas?

Lynch: They are quite fundamental. We're going to be facing an enormous amount of observational data, of experimental data, of data that represents the results of various kinds of simulations. Software is, in itself, going to become an important component of scholarship and scholarly communication, and

frankly, I don't think we're well-equipped to manage this—at many different levels. Those who are starting to think about this issue talk about a new breed of support people called *data scientists*. But it's not at all clear where we're going to find these people, how we're going to train them, what they need to know, or where they fit in our organizations. Obviously, libraries are going to have a growing role in this issue, but to meet some of these needs, librarians are going to need much deeper disciplinary expertise and also information technology expertise than they typically have.

We have many structural and funding problems. Our system for funding research tends to focus on relatively short-term grants to faculty and hates to invest in long-term infrastructure. Yet a good number of grants today produce volumes of digital material that need to outlive the life of the grant, often by decades. We have no funding strategy for these needs. By default, in most cases, the institutions that host the research are having to deal with

the ongoing responsibilities and the associated costs. Compared with other nations, the United States has a much more limited system of disciplinary data archives and data management. A few specific areas are very well supported—for example, disciplines within the scope of the National Library of Medicine and its National Center for Biotechnology Information, or planetary exploration data. But there are huge sectors with little or no support on a national scale. So the task is falling to our colleges and universities. Institutions are very nervous about the funding implications—and rightfully so. And I want to be clear that this is not a problem only in the sciences; it's coming up in the humanities and social sciences as well.

Again, there are intriguing questions surrounding how we make decisions about how long we need to keep data. Let me give an example. Suppose you have a project that produces a set of databases. These databases may be of interest to other researchers in the discipline, and you could argue that a disciplinary peer



“We’re recognizing the deepening interdependence between research officers and both information technologists and librarians as we deal with changing patterns of scientific practice.”

review should determine where investments should be made for retaining such data sets. In theory, you could imagine a science funding agency conducting a disciplinary-community-driven, peer-review-based process for allocating resources to support long-term data management (though, as I described earlier, that isn’t the case today—mostly, this responsibility falls, haphazardly, to individual host institutions). But this won’t be good enough. We’re starting to realize that more and more data has interdisciplinary value and that it may be reused, especially over time, by researchers in pursuits quite different from those of the investigator who first produced the data. Thus, leaving that conversation within a specific discipline may be quite inadequate. Historically, libraries have adjudicated some

of these issues and have facilitated the multidisciplinary use of published information. We’re going to need to figure out how to do the same thing with enormous collections of data resources—collections that don’t necessarily come through traditional publication mechanisms and for which research libraries are just beginning to sort out their roles, responsibilities, and relationships.

Hawkins: Has the role of CNI changed over the past decade?

Lynch: Yes. Let me give you an example. When we first started CNI, one of the fundamental ideas, one that I think remains absolutely valid to this day, was the notion of engaging librarians and information technologists together to address the op-

portunities and problems of this new environment. That continues to be a fundamental part of CNI’s role, but the scope of the collaboration necessary to address the opportunities has become much broader and more complicated. Librarians and information technologists are essential to this discussion, but I suspect that they no longer represent the full range of skills and knowledge necessary to grapple with many of these opportunities. If you look at CNI today, you’ll see that we’re increasingly reaching out to people involved in teaching and learning, to faculty, to those engaged in other parts of the scholarly communication chain. We’re recognizing the deepening interdependence between research officers and both information technologists and librarians as we deal with changing patterns of scientific

practice. The scope of engagement has become much richer over the years.

Hawkins: As you noted, CNI began as a sort of interjection or bridge between library organizations and IT organizations. What's your take on the current state of interaction and collaboration between libraries and IT units on campuses?

Lynch: I think it's certainly a lot better than it was at the time CNI was launched. If you look at a strategic and senior-management level at most campuses, both the chief librarian and the CIO know that they need to talk to each other frequently, they need to collaborate, and they need to work together within the broader campus structure to deal with various issues. For example, the idea of institutional repositories has widespread implications across campus. From the general counsel, to records management, to the faculty, to students, to the chief research officer, to the outreach PR and public engagement people—they all have a vested interest in aspects of the insti-

tutional repository work. Yet when we looked at this in a study we conducted on institutional repositories, it was clear that an alliance between the library and the IT groups often led the way.

Having said that, I think there are still difficult problems to work out at the on-the-ground level between the library and IT organizations. The two groups have somewhat different cultures, and although it may be relatively easy to talk about the policy behind collaboration, actually making the collaboration happen remains fairly challenging. I'm encouraged by various developments over the past few years—for instance, the fact that you now see people who have followed career paths that have moved back and forth between the library and the IT groups in an institution. That's one of the best ways to further effective collaboration between cultures. Programs like the Frye Institute have made a big difference in how those in the new generation of leaders understand what's going on in both areas and in their ability to work across them. So, I see a lot of very promising signs.

But there is also a new, third group emerging: people involved in informatics. As we look at faculty and student needs, we are increasingly discovering a growing demand for what we at CNI are characterizing—for want of a better phrase—as *informatics support*. This has some elements of IT knowledge, some elements of library and information science expertise, and some elements of data management skills; some disciplinary knowledge or context is also very helpful. This calls for a team approach including both libraries and IT organizations but also going beyond both at some level. It's unclear to me where we're going to find these people in sufficient numbers and how they're going to fit into the organization. But there is clearly a growing demand for this kind of support. Addressing this need is going to be a new and challenging field of collaboration between the library and the IT groups. *e*

Part 2 of this interview will be published in the May/June 2006 issue of *EDUCAUSE Review*.