Clifford Lynch, director of the Coalition of Networked Information, told the final plenary session of the ASIS&T 2007 Annual Meeting that we are confronted with a confusing situation, with a variety of institutions that have served us well now facing “dysfunctional” economics. The former ASIS&T president says that in periods of confusion it is sometimes useful to go back to first principles about what we’re doing. As an example, he referred to a study from Ithaka, a not-for-profit organization spun off from the Mellon Foundation. The report considers university publishing in a digital world and says that the future of university presses is not encouraging. But, Lynch says, he was struck by the report’s then asking what needs to be done to secure the future of the university presses and their services. The first question, to him, is rather to consider what we were trying to do when university presses were set up. Are university presses the best way to accomplish those goals today?

Before anything else, though, we need to confirm the basic goals. Is communicating scholarship part of the fundamental mission of universities? Lynch believes this to be true, but thinks that if you polled the leadership of a good selection of U.S. universities, you’d get a range of responses, some of them quite equivocal or lukewarm, to that question. Universities must resolve fundamental questions about the extent to which their mission encompasses the dissemination of scholarly work. Focusing on that question will help drive the discussion of the future of university presses – and many other policy and resource allocation choices facing universities today.

Lynch told his audience he also wanted to highlight another question: Do universities still have a fundamental role in stewardship of the intellectual record? That role is a big part of what university libraries do. But those libraries are facing enormous challenges in a world where scholarship is becoming more data intensive. They are being asked to underwrite preservation and become more involved in things like institutional repositories. Universities also support archives and museums that do this kind of work. Another problem is that the materials held by our cultural institutions are changing in a fundamental way.

Lynch says museums have a tradition of preserving authentic materials. Seeing a 1000-year-old artifact inspires you in ways that a reproduction cannot. But our ability to create digital surrogates is getting very good. We can now create digital surrogates that are good enough to satisfy most scholarly and even recreational uses of the material. He has been impressed by the very high resolution of the Digital Michelangelo Project (http://graphics.stanford.edu/projects/mich/), a system by Marc Levoy and his students at Stanford through which you can explore Michelangelo’s David in three dimensions. Lynch suggests that part of the stewardship of rare materials probably ought to be getting the surrogates out there, in case of disaster overtaking the originals.

It is clear, Lynch says, that data curation is going to be a big issue, not only for marquee science projects, but also for smaller scale scholarly efforts across all disciplines – humanities, sciences and social sciences. If you are running a large hadron collider, you are supported by people with lots of expertise and a dedicated staff. But if you’re an individual humanist, even at a large research university, you may not get any grants for digital work. Perhaps you get some support from your library or your IT department. We
face a real challenge on how we organize the work force and organizational support units within our universities.

In the digital world our journal delivery system is not designed to allow text mining. Text miners get into trouble because they download a lot of articles, which sets off alarms with publishers and vendors. We have some real challenges about our license agreements and how we are going to support text mining. We also have some “terribly weird and complex” intellectual property issues. There is a notion that when things pass through the human brain, sometimes they are derivative works, and sometimes they are not. Copyright law tends to assign a low place to algorithms, because of the idea that computers can’t think. So is the output from an algorithm for text mining just a derivative work? Lynch says he doesn’t know much precedent in this area, but we’re running up against a set of new challenges with very high stakes.

Consider Google’s mass digitization program. Because of copyright restrictions, Google provides only a few snippets of the material. But internally, Google has a database of the current literature on which it can compute. We have no way of knowing how much computing they’re doing and what they’ll do with the results. If they could develop the right kinds of text-mining ability, they could exploit it for purposes we could hardly imagine.

We are also seeing interaction: it’s “the coin of the realm” in Web 2.0, a term Lynch doesn’t like very much. Certainly he thinks that publicly generated content is an important development. There has been a lot of interest around social tagging. People are looking for “happy children” or “melancholy landscapes,” concepts outside traditional classification. The other side of tagging, of course, is less about assigning terminology and more about imprimatur: “I found this interesting.” This development takes us into concepts like social filtering. Consider this: As we digitize our special collections and cultural records, we’re opening them up to the public far more than ever before. The depth of description is potentially infinite, but the actual description we have is shockingly limited: “500 street scenes of Manhattan, 1951.” We make them available, and people have their own ideas for describing things: “That’s my grandfather’s fruit stand on the street scene, and that’s my grandfather standing in front of it.” There are wonderful, fascinating questions about how we structure this across organizations. But we run into these complicated problems about privacy, human subjects and copyright.

Lynch concludes that there are many more things to discuss about our relation to the cultural record in the digital environment and the stewardship of that record. He adds that he is deeply intrigued by the idea that the digital environment offers new roles, new responsibilities and new opportunities to create cultural archives.

Clifford Lynch, 2007 ASIS&T Annual Meeting plenary speaker, is executive director of the Coalition for Networked Information, 21 Dupont Circle, NW, Washington, DC 20036; email: Clifford<at>cni.org