Panel 1: Continuity and Change in University Scholarship

BERNARD FRISCHER
Institute for Advanced Technologies in the Humanities, Alderman Library, University of Virginia

JOHN UNSWORTH
Graduate School of Library and Information Science, University of Illinois, Urbana-Champaign

LORCAN DEMPSEY
Harry Ransom Humanities Research Center, University of Texas at Austin

ABSTRACT. Information technology developments over the past 3 decades have profoundly influenced university scholarship, but the relationship has been characterized by aspects of both continuity and change. This panel of presentations and ensuing discussion consider ways in which these developments drive changes in scholarly practice and present challenges for research libraries on issues such as interoperability or facilitating access in a networked environment.

KEYWORDS digital scholarship, interoperability, disclosure, digital workflows, scalability

TOM STALEY

Our distinguished panelists will address issues related to research and scholarship in the academic disciplines. Specifically, they will discuss not only how the academic disciplines are constantly reshaping themselves but also how new dimensions of information delivery and digitization change our structure of knowledge and how research libraries are responding, or should

Address correspondence to Bernard Frischer, Institute for Advanced Technologies in the Humanities, Alderman Library, University of Virginia, P.O. Box 400115, Charlottesville, VA, USA22904-4115. E-mail: bernard.frischer@gmail.com
be responding, to this changing environment in which the world’s texts are being electronically copied, digitized, and linked.

Thomas Carlyle spoke of a simpler time when he wrote in the 19th century, “The true university of these days is a collection of books.” Jim Duderstadt spoke of the metauniversity in a global universe. I doubt that Carlyle would be able to find it.

Our three panelists, each in his own way, have already contributed to this dialogue. Bernard Frischer is the current director of the Institute for Advanced Technologies in the Humanities at the University of Virginia. He is at the forefront of scholarship in the application of digital technologies to humanities research and education. He founded and directed the Cultural Virtual Reality Laboratory at the University of California, Los Angeles, where he oversaw such projects as the virtual recreations of the Roman Coliseum and the Roman Forum.

John Unsworth is dean of the Graduate School of Library and Information Science at the University of Illinois, Urbana-Champaign. He was the first director of the Institute for Advanced Technologies in the Humanities at the University of Virginia. More recently, he served as chair of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences, which explored the high-performance computing tools available to researchers in these fields.

Lorcan Dempsey is vice president for research and chief strategist at the Online Computer Library Center, better known to most as OCLC. He has written and spoken widely on resource discovery services, metadata, the distribution of information and library systems, the changing position of libraries, and the curatorial traditions of libraries, archives, and museums.

BERNARD FRISCHER

Thanks for that nice introduction, and thanks for the invitation to speak to you today about how the humanities have been affected by information technology in the past 20 or 30 years since I was hired at the University of California, Los Angeles (UCLA), as an assistant professor of classics.

Does it matter that, in general, scholars in the humanities have traded in their typewriters for a personal computer? I think that we would all answer that, yes, it obviously does matter. The medium may not be the message, but the medium certainly does change the ways in which an ever-growing number of scholars craft and communicate their messages. The further back we go down the career path from full professors to assistant professors to K–12 students, the more we can see these changes at work.

Even though the famous physicist, Neils Bohr, usually is correct in claiming that predictions are very hard to make, especially with regard to the future, in this case he isn’t. We can see the future if we just look at the world
First, looking at the end of the scholarly communications pipeline, one striking difference between many works of scholarship today and those of 30 years ago is not only the vastly increased number of digital publications but also the marked increase of scholarship in the humanities that is coauthored. Coauthored publications grow out of collaborative online communities, like MySpace for kids or the one created by Professor David Germano for the study of Tibetan culture. His Tibetan and Himalayan Digital Library integrates the efforts of nearly 1,000 scholars around the world, and it fills more than 10 terabytes of storage on our library’s server.

Another example is Raphael Finkel’s Suda On Line project, founded in 1998. The Suda is a very important 10th-century Byzantine encyclopedia of classical culture. Its 30,000 articles had never before been translated into English, despite its precious biographical information about Greek writers such as Hippocrates, Sophocles, and Plato. Under Professor Finkel’s leadership, almost 20,000 articles have now been translated and published on the Internet since 1998.

If we compare the Internet to the U.S. Post Office, which is what we would have had to use 30 years ago for such a collaborative project, we can see that online collaborative research is obviously infinitely more feasible in terms of cost, speed, quantity, and especially the quality of interaction that it affords scholars. Something like the Suda project or the Tibetan and Himalayan Digital Library didn’t exist 30 years ago because, for all practical purposes, it would have been impossible for them to exist.

Another thing that couldn’t exist, mainly because it would have been prohibitively expensive, is the large-scale publication of materials typically housed in the special collections department of an American research library. I’m thinking of maps, medieval manuscripts, ancient papyri and incunabula, and so on. In recent years, collaborative projects by scholars and librarians have proliferated, providing comprehensive free access to such materials.

One of many that I could cite is the incredibly useful online Perry-Castañeda Library map collection at the University of Texas (UT), a collection with more than 5,000 digital maps online and available to the public and 250,000 available here at the university—an extremely useful collection.

Finally, let me also mention at least one example of something that could not have existed 30 years ago because it could not have been supported by any technology available at the time—a fully interactive three-dimensional model of an important historical site that no longer exists, such as the Forum or Coliseum in Rome, the Acropolis in Athens, or the Temple Mount in Jerusalem. In the past 10 years, a whole new field of digital archeology has sprung into existence, providing us with such recreations.

This panel concerns continuity as well as change, so let me mention something that has remained pretty much the same—the need for peer
review. Behind this need, there are some practical reasons, including the issue of what counts toward promotion and tenure, as well as good, normative motivation. Peer-reviewed work is almost always better than work that isn’t.

Here, though, we can see an important way in which the digital revolution has been getting ahead of itself. For some years, scholars—including myself—have, in effect, used the Internet as one great big vanity press, releasing their work to all and sundry without the benefit of peer review. I think that the reason for this is obvious. It was not that we didn't appreciate the virtues of peer review or that we didn’t want our work to be judged by our colleagues but, rather, that there was no mechanism whereby we could submit our digital creations to them for comment, acceptance, or rejection.

This is gradually changing. The Suda project is a case in point. Here, the peer-review process is built into the collaborative publication pipeline itself. In this case, the similarity to what we were familiar with in the Gutenberg age also hides a difference. In the archaic world of print publication, peer review had to occur prior to publication, for obvious practical reasons. In today’s digital age, a translation submitted to the Suda project can be published before it is peer reviewed, and that’s how they do it.

We clearly need many more peer-reviewed online publications like the Suda project. This need is starting to be felt across the humanities. For example, a recent issue of the Chronicle of Higher Education announced that the Journal of the Society of Architectural Historians was going to start an online version, a version that Editor Hilary Ballon assures us will adhere to the same high qualities as does the print journal.

Let me conclude by giving a glimpse of the near future. The field of classics was an early adopter of information technology and now faces the ironic problem of managing success. There’s an embarrassment of online riches available to classicists, but each Web project has its own peculiar features and way of doing things. The typical resource—for example, a collection of Greek or Latin texts—does not communicate automatically with other online resources, such as a Greek or Latin dictionary or an online bibliography.

The answer to this problem inevitably lies in standards and in the interoperability that standards make possible. In the achievement of interoperability among our online resources, librarians and information scientists can help, and I’m happy to report that they are helping by teaching us how to exploit such things as the Fedora project, with its flexible and extensible digital object and repository architecture.

The field of archeology lags behind classics, but recent developments in professional organizations suggest that this is about to change as archeologists realize that information technology offers them better ways to do their traditional tasks. Given their role as our primary conservators of the human record, it is ironic that the major challenge facing archeologists today is how to preserve their own digital data. Here again, as the DSpace
Federation shows, the librarians are ahead of the scholars and are adapting their traditional role of storing manuscripts and printed works to the needs of the digital age.

This brings me to my conclusion and a final point of comparison between the situation today and that of 30 years ago. I refer to the relationship between librarians and scholars more generally. Three decades ago, I would say that the relationship was rather standoffish. As a professor at UCLA, it never occurred to me that I might want to use the library as a place to write and actually produce my scholarship. It was just a place into which I made periodic forays to find, or not find, a journal or book that I needed. I think that my experience was typical of many other humanists. There was no particular reason to develop a close, working relationship with our local university library, and it never crossed our minds to ask for space in the library, except occasionally for a humble carrel.

Today, the situation is quite different. Because of the wealth of online materials, many scholars can do their research and writing anywhere; yet, ironically, we find ourselves going to the library no less frequently than we did before. We go for help with using online resources, to learn about software that supports our scholarship, and, at least at my university, to work in a research unit called the Institute for Advanced Technology in the Humanities (IATH). At IATH, scholars have permanently been allocated more than 2,000 square feet of space in the main library, to produce their digital work and to realize their digital dreams in collaboration with information scientists and the technical staff who support them and their collaborative research projects.

So, the research library is still just that—a place to promote and support the research of the faculty—but the way that it is fulfilling its mission has begun to change in some important ways. Projecting forward of current trends, I therefore will end by predicting a rosy future for the university library, which, with or without books within its brick-and-mortar walls, will still be the place to go on campus to find the key resources that we humanists need to do innovative scholarship.

JOHN UNSWORTH

At the moment, we know a lot more about our digital collections in the research library than we know about what people do with those collections or wish to do with those collections. To illustrate the situation, I’m going to use an example that’s close to hand, from a project on which I’m currently working, as a way of explaining why I say this. The name of the project, NORA, is an acronym that stands for No One Remembers Acronyms. That’s how you can remember it, see?
NORA is a text-mining project that works on humanities content in digital libraries. We’re looking at 18th- and 19th-century novels in English that have been contributed from library collections and from scholarly projects. We’re taking the texts that are already marked up by the libraries and by these scholarly projects and trying to bring them together for text mining.

You need to gather material from across collections in order to put together a coherent and large enough set of material to be interesting. One collection doesn’t have everything in a certain area. You also need to bring these things together for practical reasons, in order to preprocess them together and in order to index them together. When you start to ask for statistical information about these collections, that statistical information is meaningful only if you’re asking about it in the context of everything else. You can’t get inverse document-frequency counts of, for example, certain pieces of vocabulary in one document and then in one collection and then add them to the numbers for the same thing from some other context and get a meaningful answer. You actually need all that stuff together.

The resources that you find, as soon as you look at them, are clearly expected to be used in situ, in their original context. They had no idea—it’s like they woke up in a bad dream, with no clothes on in a public place—that they might be called on to be used and in which use meant to be browsed or possibly searched. We need to preprocess those to get collection-level contextual information, but they don’t anticipate that kind of preprocessing.

In the NORA project, we need to be able to give someone at least a bit of context to read once they get to a data point—once they’ve drilled down to something to say, “Wow, that’s an interesting point; where does that come from? Show me a few words of context.” I need to be able to show that context to them even if they don’t have rights to whole documents. I need to be able to take someone from that snippet to the original publicly disseminated address of the whole document, but, again, the whole document doesn’t carry that original address with it. It doesn’t have it anywhere in its public address, so there’s no way to do that.

I need an environment that allows users of these collections to contribute their results and share them with one another. I need an environment that allows users to share intermediate artifacts in the research process—that is, the preprocessed collections that they developed for their particular uses, because that preprocessing is what takes all the time and the actual analysis takes relatively little time. So, if we suppose that two people might be interested in the same subset of content, there would be value in keeping that subset around once it had been preprocessed.

For this project, I need an environment that allows users to enrich and improve the data that they’re analyzing with user-supplied metadata, where that metadata might be anything from normalized spelling to suggested corrections of underlying content or just richer thematic information about the
document. But, they need to be able to do that without undermining the integrity of the original collection.

I’m going to need all of that, incidentally, from publishers if the libraries are not the owners and disseminators of the texts. If the library’s just a gateway to the publisher, then I need this from the publisher, not just from the library.

So, what’s the problem? Well, texts that are prepared in order to represent an original—when that’s the motivation for doing the digitization—often make it difficult to collate and compare and analyze. I’ll give you a very simple example, which is line-end hyphenation. Faithful encoders tend to encode line-end hyphenation when they hit it, when they’re encoding a printed text. Textual editors probably will thank them for doing that, because this is a potentially useful piece of information for textual editing, but line-end hyphenation makes it very difficult to know that this word that’s hyphenated is the same as the unhyphenated form of the word. It’s not impossible, but it’s a little bump in the road.

Texts prepared under the assumption that they always will be used in the same way, for browsing and searching in the same environment in which they originally were prepared for publication, have a tendency to leave certain kinds of information implicit. It’s implicit somewhere else in the system. It’s not in the document itself. The problem of being able to return to the publicly disseminated version of the document is just one instance of this, but there are lots of other ways in which this happens.

Texts for which you expect to have use restrictions taken care of at a higher level, before you get to the texts, often don’t carry the conditions of use with them; if they did, we wouldn’t really have a framework for resolving the relationship between the context of use at the moment and the conditions of acceptable use for that particular document.

In most cases, even though the underlying source of these documents is XML, there’s no direct way to get to that underlying source. I’m usually getting a rendered HTML version, with rare exceptions. We got our documents for the test bed for the NORA project by mailing thumb drives to some of my friends in the library and scholarly community; they loaded the source onto the thumb drive and mailed the thumb drive back. That worked pretty well, but it was a little slow.

Where would the processing and indexing of these scattered collections gathered for a particular research project get done? Would it be on the home server of one of those collections? Would it be on the end user’s machine? If it were on the end user’s machine, how would those intermediate research artifacts get stored and saved?

How would someone who wanted to verify a colleague’s results reassemble the original data and retrace the processing? It’s like the citation problem, with which we’re all familiar, but it’s not just how do I cite my
results; it’s like citation with provenance and an audit trail attached. It’s a bigger problem.

What’s the solution to this kind of problem? Well, one solution would be to run screaming in the other direction when someone comes to you with a problem like this and say, “You must be the only person who wants to do that, and it’s not worth solving it for you.” At the moment, you could probably get away with that, but I don’t think that you’ll be able to get away with it for very much longer.

So, for starters, a more constructive solution would be to support the data-standards communities, and I agree with Bernard Frischer that standards are part of the answer; but they’re only part of the answer. I also think that those communities need to include users of data, as well as curators and technical experts, in the development of standards. You could expand on experiments like DLF Aquifer, by expanding them from being about collections sharing to being about shared collection processing, not just federated searching.

But, how could you actually do other things with subcollections drawn from participating libraries? You could support explorations like the Pathways Project and other efforts to build frameworks that envision the use of digital content in the context of scholarship in which the data or data streams come from lots of different sources and in which people do arbitrary and somehow unpredictable things that still need to be recorded and still need to be traceable.

In general, I think that we need more conversations, projects, committees, and communications that involve particular domain scholars, particular collection experts, and relevant technical experts focused on some kind of processing of data beyond searching and browsing. I also think that, last but not least, we need more systematic studies of the uses to which scholars and scientists put information in digital form—studies of the tools that they use and how they use them and of the tools that they need and why they need them.

Some of those studies have been done. I certainly have people on the faculty of the Graduate School of Library and Information Science, such as Carole Palmer, who work on this. There are people at other schools, here at UT, at the University of Michigan, and at other places who do those kinds of studies, and I think that increased collaboration between the library community and people interested in those questions would be good. A lot of that work also goes on at the OCLC.

LORCAN DEMPSEY

The earlier mention of Thomas Carlyle is interesting. Thomas Carlyle is the archetypal user who knows better. When Anthony Panizzi, one of the
founders of modern library cataloging, managed the British Museum Library (the precursor of the British Library), he was developing his rules for a British Museum catalog. Carlyle, who was on the committee to which Panizzi reported, was scathing about the obfuscation and complication of these cataloging rules. To Carlyle, it was quite simple—you just do a list of the books, and then you can find them afterward. So, Carlyle went on to be one of the founders of the London Library, a private circulating library that exists to this day, but I always think of him as the library user who knew better.

I’m going to talk a bit about libraries with respect to three subjects, each of which is systemic in nature and grounded in the historical reality that we have a set of institutions that developed around physical collections and that are vertically integrated around those collections. Many of the difficulties that we face at the moment—or, if you prefer, many of the challenges or opportunities—relate to the fact that we now live in a world that is constructed and organized in a different way. The network is the center rather than any physical place.

I want to speak of three things: workflows (and network flows), the scholarly record (it isn’t what it used to be), and access to scale, which picks up on some of the comments at the end of Clifford Lynch’s session and moves to the network level.

Workflows

If we think about the library historically, at one stage people had to build their workflow around the library. That tended to happen in a physical world in which people built their workflow around particular elements of specific collections. So, people built their workflow around the library—around what was available.

What’s happening now is that people are building their workflow in networked environments. You have prefabricated workflows such as the course-management system, the institutional repository, and electronic lab books. You have a variety of workflows that are constructed, and a lot of discussion about how helpful they are.

You also have practices developing on the network that are constructing digital identities, lifestyles, or ways of working around things as simple as RSS aggregators, My Yahoo, and other similar services. In the near future, there will be a whole host of tools to help you build workflow in a sort of self-constructed way, because of the centrality of the network. Part of the issue for libraries, then, is that we still operate on the assumption that people build their workflows around the library, rather than imagining how we might make materials and services available in workflows that have moved elsewhere.

With regard to John Unsworth’s comments, for example, we historically have focused on discovery, saying come here and look for resources. I think
that, increasingly, we’re going to have to focus on disclosure and dispersal. We’re going to have to discover ways of disclosing resources to workflows. That might mean exposing things to Google and the search engines, and it might mean exposing things in ways that would allow them to be consumed by RSS aggregators or easily plugged into course-management systems and a variety of other things beyond that. The onus will be on making resources available in such a way that they can be disclosed or exposed into these other workflows, rather than expecting people to come and discover them.

Dispersal relates, quite strongly I think, to what John was just saying. There’s a phrase by Raymond Yee of UC Berkeley, which is quite nice: “gather, create, and share.” People want to gather stuff, create stuff, share stuff. At the moment, our resources don’t play very well in that context.

In the context of people wanting to acquire materials, manipulate them, use them, and have them placed in ways that they can be encountered within workflows, what we’re seeing is a general move away from the idea of managing content toward the idea of managing consumption. What becomes important is thinking about how people want to use resources and then thinking about what is required in order to facilitate that use. Again, that’s not something we’re used to thinking about in the physical world, because, in that context, people take away the materials and use them in their own environments. In the networked world, there has been a collapse between those different environments, and managing consumption has become as important as managing content.

The Scholarly Record

It might be fair to say that, historically—and I’m sure various people in the audience will resist this—publishers created the scholarly record through their selection activity and their editorial activity. Libraries, in turn, selected from a preestablished scholarly record. The role of libraries was to select the materials that moved into libraries and then were preserved and managed in that context. One could equate, approximately, the published record—published as it was then understood—with the scholarly record.

Now, of course, that is no longer the case, and Clifford Lynch described various reasons why this is so. If you think about the types of material that have been discussed already at this meeting, you have data from a variety of sources. You have Web sites. There was some discussion previously about collecting Web sites in the context of particular scholarly endeavors, of institutional records, and of area studies. You also have a whole range of materials that are not conventionally published. This means that what becomes part of the scholarly record and what is going to be managed as the scholarly record are no longer straightforward.
This creates collection-development issues for libraries, about which we don’t yet have a consensus. It creates issues in terms of where to invest resources over time and in terms of thinking about what constitutes the scholarly record of the future. This, again, relates to uncertainty about behaviors and requirements. What will the scholar of 2050 expect you to have collected?

With research libraries taking a more active role in managing the parts of the scholarly record that are not conventionally published, as well as those that are, and in acknowledging the blurring that happens, there will have to be a much stronger move toward a more archival perspective. Issues of provenance and authenticity (John mentioned audit trails)—the context from which particular parts of the scholarly record emerge—become increasingly important. The uses to which it’s put become important, and a really critical issue—one on which we don’t have a good handle—is citability.

Citation is fundamental to scholarship, but it’s unclear how one cites some of these things with the confidence that one is citing the thing that was actually the object of investigation or analysis. At the moment, we have that issue in a wholesale way with Web sites: People are citing things that are no longer there or for which the context has changed. If you look at some of the work that’s being done in an e-science context, one of the big issues is how does one cite; when you want to cite a curated database, for example, you want somebody else to be able to recover it as it was constituted at the time that it was cited.

Related to citation is what I somewhat glibly like to call “ex-citability,” or “ex-citation,” which means that you want to cite things that are executable. So, you have a mixture of executable and citable. There’s a whole set of issues about citation and citation fundamental to scholarship. This isn’t a question about persistent identification. It’s more than just persistent identification. There’s a whole set of issues, I think, about the scholarly record—both discretionary issues about what one cites as part of that scholarly record and technical and professional issues about how one manages it.

Access to Scale

The range of what the library has to do in the context of the environment discussed yesterday by my colleagues grows and grows. It’s increasingly inevitable that the library will want to focus on where it creates a distinctive impact.

Those things that are routine or done across many libraries should be moved to the network level—either moved to a shared place or collaboratively sourced in some other way. Network-level sourcing allows libraries to address several issues, including the opportunity cost of continuing to do those things, the potential for system-wide efficiencies, and the question of
how one creates impact in areas such as workflow at a level that’s difficult to do within an individual institution.

Thinking of some examples from a discovery point of view, I mentioned earlier that one of the issues is a move from discovery to disclosure. Discovery, though, may happen at a higher level. If you want to be able to disclose to discovery environments like Google Scholar or if you want to make sure that your materials are available to people through Google Scholar, then that effort may become as important, or even more important, than what you do locally. There’s an access-to-scale issue. What does it mean to engage with discovery environments that are above the level of individual institutions (from an OCLC point of view, WorldCat.org is an example), and how does that then relate to disclosure in other environments?

Another issue is longer term opportunity costs in the management of print collections. There’s a lot of discussion about the extent to which a library’s distinctive impact is bound to the richness and depth of its collections. Over time, how much of that value remains in that particular configuration? To what extent will that value be moved to more shared contexts in the context of mass digitization, shared storage initiatives, and the thinking about how one divides particular subject areas?

In time, I think that our view of collections is going to change. Groups of libraries will develop a much more shared, or collective, view of aspects of those collections and will think about how to deal with the opportunity costs of managing large physical collections, given questions about where value resides and where it must be.

Preservation is an obvious example, and archiving the Web is another obvious example. Again, these are issues about which one wonders whether they’re best tackled at the institutional level or at another level. Things that you might not normally think about—for example, social networking approaches such as tagging, recommending, and collecting reviews, with which various libraries are experimenting—also will need to be examined for scale. These have to have a local presence, but, really, does any one library have the critical mass to be able to do significant recommending based on tagging or reviews? That really needs to happen at a higher level, and then scale needs to be made available to libraries. By a higher level, I mean that it needs to be aggregated and consolidated in some way that there’s a level of participation in a scale activity.

Usage data is another such area. I mentioned earlier the idea of moving from managing content to managing consumption, a phrase that I picked up from the search engine FAST and some of its work. As I said, it represents a general perspective.

One of the things that we don’t do very well at the moment is aggregate usage behavioral choice data, which, in turn, can refine services. This relates to understanding what people are doing. Again, if you want to have good management intelligence, if you want to have good recommender
systems, and if you want to do other things, that type of data becomes very important. What people are using, what choices they’re making, and what people are buying. We don’t have good ways of aggregating that data and making it available.

There are a variety of areas in which we increasingly want to do things in an aggregate and shared way. We want to collaboratively source activities. I think that, given the complexity of the environment and some of the challenges, we need to become much more purposeful—that is, have a much more instrumental attitude toward various organizations that are available or create new ones—in order to make sure that we do locally what’s important and that we find good ways of doing what’s important in a shared way.

QUESTIONS AND ANSWERS FOLLOWING PANEL 1

Peer Review

The discussion following the early panel was rich and flowing, covering a wide range of topics. A senior professor of liberal arts at UT made the observation that the increasing prevalence of information technologies raises the standards of scholarly competence beyond the corpus of the discipline in which a scholar had been raised. Referring to the work of Mike Davis, she argued that postcolonial scholars’ arguments would be judged by the ability to step beyond resources familiar to the discipline and to avail themselves of metrological, epidemiological, and demographic material. She suggested that librarians could become more important in pointing researchers toward collections and resources that could lie beyond their ken.

In response to her comments, John Unsworth observed that scholarly inertia can blunt the speed with which such a shift in accountability could take place. Should it occur, he said, it will be the discipline that directs the shift in accountability, not the librarians. Bernard Frischer agreed with his fellow panelist’s take on scholarly inertia but suggested that the role of the librarian in the scholar’s life was becoming more important.

Frischer also observed that, although peer review remained an essential part of the scholarly communication process, technological developments had reversed its place in the scheme of things and perhaps made it more important. For the process to work in the print world, peer review had to take place before a work was printed, for the most part. In the digital world, peer assessment can follow publication; teams of authors will become larger, interdisciplinary works will emerge in greater numbers, and the scholarly community as a whole will grow.

Transformative Changes

Another participant raised the question of whether the advancements in information technologies have resulted in transformative scholarly projects.
Unsworth observed that he felt that forums such as the work of the American Council of Learned Societies Commission on Cyberinfrastructure in the Humanities had fundamentally transformed the dialogue. That change, he suggested, is permeating the scholarly societies and the college faculties. “I see more people being hired to do digital work and more people being tenured for doing it.” He felt that, although gradual, the change was not only under way but was irreversible.

Bernard Frischer agreed and pointed to the National Endowment for the Humanities digital humanities initiative as a case in point. He recalled an observation from that agency, some years prior, that his own work in digitally reconstructing Rome was not scholarship. His grant success in recent years is one indication that perceptions had indeed changed.

Scalability

The conversation between the panelists and the audience turned from the concept of transformation to that of scalability. Bernard Frischer pointed out that digital tools such as Suda On Line: Byzantine Lexicography enabled scholars to track every contribution to the discipline, making it possible in a peer-review sense, for purposes of tenure or promotion, to be very precise about the quality of a scholar’s work.

Issues regarding the vast scale of the digital universe and the travails of text mining then became the focus of discussion. To Lorcan Dempsey’s hypothetical problem of mastering the body of knowledge that amasses in a “thousand blogs on Shakespeare,” Bernard Frischer suggested that text mining was the answer.

From the audience, a technologist offered an interesting observation about scholarship in a technological age. It is far easier today to search a billion documents than it was a decade ago, he observed, but it is probably no easier to maintain relationships with 50 or so colleagues than it was then. Trusted peers remain the chosen source of recommendations, over people who are unknown. It also is true, he offered, that it is probably not possible to read much more than one did a decade prior.

As for the librarian’s concern with the rising cost of publication, he observed that, from his perspective, the real cost of making information available to people has become cheaper in the digital age. “What’s more expensive is figuring out what to read.” The publishers pass on to the end user the cost of enabling the latter to make that decision about what to read.

Lorcan Dempsey responded that the observations drive home the need to answer the sustainability of workflows in an expanding information universe. He pointed out that previously, in the print world, information resources were relatively scarce and attention was abundant. Now, in the midst of overabundant information, it is attention that is in deficit. The question becomes one of managing consumption.
Authority
To the issue of scale, Bernard Frischer inserted the idea of authority. At the University of Virginia, he observed, there are formative efforts to develop collaborative filtering tools and to attach them to online scholarship. The purpose of the tools, he said, would enable readers to track the commentaries of recognized authorities in a field. The use of online resources by authorities could be tracked and reported to a central database, and it would be possible for others to access a kind of online bibliography of what thought leaders in the discipline were accessing or reading.

Responding to the question of authority, a university press director in the audience observed that university presses indeed served that role in the past. However, the question has become much larger in the exploding information universe, and she remained unsure of how it will be adequately addressed. Another administrator to whom the university press directly reports added that she felt that the press still had a role in emerging disciplines by providing a mechanism for the organization of inquiry and research around journals and monographs, which enables focus.

Persistence
Presses continue to play a role in the persistence of scholarship as well, the administrator continued. We still have not answered the nagging issue of the longevity of born-digital information, she continued, and for that reason the presses continue to play a vital role in the archiving of information. We can be sure that information published in a book or journal will persist for hundreds of years; we have no such assurance about Web data.

A library director underscored the need for the persistence of born-digital information when he pointed out that both federal and private foundations are beginning to realize the need to archive the results of funded research and are turning to librarians to help craft the solutions. One of the solutions, he added, that is growing out of the work of the Section 108 Copyright Study Group would be to recognize in public law that certain institutions should have defined authority to collect and archive Web sites and that there has been no concerted effort to date, because of the lack of a public policy framework.

Copyright
A copyright attorney in the audience asked if the panel would agree that, in a rapidly evolving digital environment, some would say that copyright is a hindrance and that publishers are caught in the cross fire. John Unsworth replied that copyright is a social issue. He added that in some instances, such
as when the desire is to aggregate material across collections for purposes of study or analysis and some of that material is either licensed or copyrighted, an element of complexity is introduced that would not otherwise be there. It is, however, a social issue that can be addressed and solved. He looks forward, he said, to a resolution that advances the principle of fair use and allows researchers to look at portions of text even before the determination of whether access to the entire text is allowed.

Tom Staley thanked the panel and the audience for their participation, and the session was ended.

REFERENCE
