Planning for Digitization: Overview of Major Issues

by

Donald J. Waters

Program Officer, Scholarly Communications

The Andrew W. Mellon Foundation

February 4, 2005

THIS IS A WORK IN PROGRESS

Introduction

Good evening, ladies and gentlemen. Welcome to Charleston and to this workshop. Thank you, David, for the kind introduction, and to the organizers of this meeting for kindly assigning me the task of providing an overview of issues related to digital asset management.

The speed with which this workshop became fully subscribed after its announcement is a measure of the importance to you of the topics that will occupy us over the next two days – and perhaps of the difficult challenges that those topics represent. It may be also be measure of the frustration that we all feel in trying to manage digital assets prudently and well at a time of change when all the parts seem to be moving at once – and we are making up many of the parts as we go along.

Despite the rapid pace of change, there is one point of constancy and common ground that I should emphasize at the outset of this workshop. That point is our common mission in the service of the academy. Let's be crystal clear that the digital assets of which we speak tonight and in the next few days are resources for research and teaching, and that our purpose in managing them is to advance knowledge and improve education. The touchstone question for all the topics that we will cover this weekend must be: how will this system or that feature advance scholarship?

Inquiring minds often want to know the criteria or other features of the Mellon Foundation's digitization initiatives. My reply is always that Mellon has no digitization initiatives. We do, however, have a variety of programs to help advance scholarship that may involve digitization and other uses of digital technologies. So with respect to the management of digital assets, we must remember – to paraphrase the slogan from the 1992 Clinton campaign – that "It's the scholarship, stupid!" That should be our talisman this weekend. Write it down on the back of your hand. Raise that hand if the discussions become at any point too abstract or start veering off into technical or other kinds of esoterica.

But do also remember that these are complicated issues that require creativity and imagination. We must teach each other and share what we are learning in an open cooperative

spirit. Our biggest digital assets are each other – and your biggest digital asset this weekend may be sitting next to or across from you. In saying this, however, I do not mean to underestimate the challenges of communication in managing the human factors of scholarly digital assets. We all depend, for example, on the genius of our programmers, but some of you are surely aware of what distinguishes an introverted programmer from an extroverted programmer. An introverted programmer stares at her shoes when she is talking to you; an extroverted programmer stares at YOUR shoes.

Talking in code is also endemic in this field and is another major obstacle to clarity and understanding. Some of you may have heard the story about the student of computer science on his first day here at the College of Charleston. He had entered a strange, bizarre world. The only sounds were the occasional calling out of a number by one of the professors, followed by laughter. One professor would say "52," there would be a short pause then peals of laughter. Someone else would say "713," same thing, everyone would fall down laughing.

"What's going on here?" the new student asked his advisor. "We're telling jokes," the advisor replied. "You see, we've all worked here so long we know each other's jokes. There are a thousand of them. So, being information theorists we applied data compression. We just assigned them all numbers, 0 through 999. It saves a lot of time and effort. Would you like to try? Just say any number from 0 to 999."

The student wasn't fully convinced, but he tried. Very quietly he whispered, "477." There was hardly a murmur. He looked at his advisor. "What's wrong?" he said. "Try again," said the advisor. So he said, louder: "318." Again, nothing; no response. "Something's wrong," the new student said. "Well," said the advisor, "it's like this: it's not so much the joke as the way you tell it!"

There is a curious sequel to this story. The student eventually succeeded by accident in the most dramatic and unexpected way. He called out a number outside the range 0 to 999. "Minus 105," he said. At first there was stunned amazement, then one professor began laughing hysterically, then another, then another, till they were all rolling about holding their sides. None of them had heard that one before.

"It's the scholarship, stupid" and clear and open channels of communication are two of the preliminary points that I want to make here this evening. The third is that the effort to ask the right questions is just as important as – and perhaps more important than – getting the right answers. I was intrigued last week by the report on MSNBC that an asteroid had been officially named for Douglas Adams, the science fiction humorist and author of the *Hitchhiker's Guide to the Galaxy*. The DouglasAdams asteroid is also known as 2001-DA-42, a striking combination of Adams's initials, the year of his death, and the famous answer to the Ultimate Question.

Readers of the *Hitchhiker's Guide* will remember that in the story:

A hyper-intelligent race of beings take the form of mice and construct Deep Thought, which was the second greatest computer of all time and space, and built to calculate the answer to the

Ultimate Question. After seven and a half million years of pondering the question, Deep Thought provides the answer: 'forty-two.'

"Forty-two!" yelled [one of the researchers]. "Is that all you've got to show for seven and a half million years' work?" "I checked it very thoroughly," replied the computer, "and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is."

Deep Thought then informed the researchers that it would design a second and greater computer, incorporating living beings as part of its computational matrix, to tell them what the Ultimate Question is. That computer was called Earth and was so big that it was often mistaken for a planet. Just as Earth was about to produce the Ultimate Question, it was destroyed, supposedly to build a hyperspace bypass. Later in the Hitchhiker series, it is revealed that it was actually a consortium of philosophers and psychiatrists that arranged to have Earth destroyed because they feared for the loss of their jobs if the meaning of life became common knowledge.[1]

If you signed up to this workshop looking for answers, there you have it. The answer to the Ultimate Question is "42." But one of the morals of the Hitchhiker story is that the certainty of answers like "42" is all but irrelevant if it is cast in the absence of a framework of meaningful, substantial questions. I take it as my job for the remaining time I have this evening to help frame this workshop with some of the strategic questions that need to be addressed in creating effectively managed scholarly resources in your institutions. The very distinguished series of speakers that the organizers have assembled will work with you over the next two days, and together you will discuss these issues in much more detail, elaborate them, suggest different emphases, and no doubt disagree with points that I will ask you to consider.

My plan is as follows: First, I want to review and critique the current debate about scholarly publishing. Then I want to recognize the big elephant lurking about in our midst – Google – and highlight how the priority of asset management questions might shift in light of recently-announced activities by Google and other commercial search engines. I will conclude by suggesting a few principles for determining when digital assets need to be managed by someone else, when collective action is needed, and when local institutions must act.

Scholarly publishing

One of the largest, most visible, and fastest growing investments in digital assets by colleges and universities over the last decade has been the purchase of licenses to electronic journals and databases. The key management task for this class of assets has been to deal with steadily increasing prices. Even before the introduction of electronic versions, journal pricing had been on a rising trajectory, especially in fields of science, technology, and medicine. As publishers have moved steadily over the last decade to add electronic versions, the pricing crisis for academic institutions has not improved but steadily worsened as provosts and librarians, especially during the booming nineties, largely agreed to finance enormous annual price increases. A provost once justified such payments to me by saying: "It is my job to feed the

[1] "The Answer to Life, the Universe, and Everything," in Wikipedia at http://en.wikipedia.org/wiki/The_Answer_to_Life,_the_Universe,_and_Everything.

hearts and minds of the faculty and staff. For healthy hearts, I must provide extraordinary annual increases in budget for health care; for their minds, I must provide similar increases in the serials budgets."

Now that resources are squeezed, attention has focused on the increasingly deleterious effects of these previous financial agreements. On many campuses, administrators have mobilized faculty, calling them to account as editors and authors of expensive journals and involving them more deeply in ongoing purchase decisions. Big deals have been undone, bundles unbundled, and there is even research underway to scrutinize publisher consolidation for cause to justify government anti-trust action. In addition, members of the academic community have embarked on a vigorous search for other, alternative business models. This search has resulted in stimulating and sometimes heated debates about the viability of a suite of options under the broad umbrella of "open access publishing."

As originally formulated, open access publishing would disrupt the current system by shifting the burden of generating revenue from the demand side through widespread use of subscriptions, to the supply side through the use of author fees, making use cost-free. Such a shift would have the benefit, in theory, of putting the principals-the scholars-in the economic driver's seat, and it would have the broad public policy benefit of lowering the economic barriers to reading and using the publications. Discussion of this idea has quickly revealed that in very few disciplines do scholars have sufficient funds from grants and other sources to pay author fees, and that there would be an administrative nightmare if academic institutions that are already financially strapped for funds were to massively reallocate budgets from library acquisitions and other sources to support author fees in any kind of fair and equitable fashion. Publications adhering to this strict definition of open access will undoubtedly continue to be created and survive, but they will probably be limited in number unless and until sources of supply-side revenue can be found that do not depend as heavily as they do now on grant support, and that do not require fundamental administrative and financial overhauls of our institutions.

In the face of these practical difficulties, the open access discussion has now morphed to focus on other ways to lower the barriers to access, such as by encouraging publishers to make articles freely accessible after a limited time during which they exploit subscription revenue, or by calling for authors to "self-archive," that is, to retain rights to make available their articles in pre-print and/or post-print form. The theory behind this approach is that if enough authors were self-archiving, then new services could arise to collect, aggregate, evaluate, and present these articles to users. At some tipping point, as yet undetermined, these services might serve to challenge and undermine the economics and inefficiencies of the current system of publication.

The key barrier to a complete transformation following this scenario, however, is precisely an asset management issue. Does it make sense for systems of knowledge to be built upon the fragile infrastructure of a network of personal Web sites that are subject to personal whims, not to mention the migratory habits of individuals? Institutional repositories might help, especially if they were to collect other related faculty output, such as underlying data and teaching materials. However, in order to justify the necessary and significant costs of such repositories, institutions like yours must develop compelling rationales for collecting, preserving, and providing access to these kinds of output. Moreover, institutional repositories must be invested

with features of the current scholarly publishing system that preserve trust in the authenticity of academic work and reliably allocate credit.

Because there is so much at stake for individual faculty in the ways that the current system confers credit and authenticity, it appears that these factors are going to be the hardest to disrupt. Even the self-archived material in physics and related disciplines in Paul Ginsparg's famous ArXive has not resulted in substantial shifts from traditional forms of publication – at least not yet. Still, efforts to build new models of scholarly communications based on institutional repositories and rights to self-archiving have gained growing interest. They remain worth exploring, and Mellon has provided a series of grants for preliminary studies of how this approach might be scaled up across disciplines.

Important as the serials crisis is, and as pregnant as the discussions about open access alternatives may be, there are even larger forces at play. These are only partially revealed in the system of scholarly publishing, and may even be obscured by a narrow focus on pricing and open access. First, whatever happens with open access, it is not likely to result in a uniform, utopian solution. As Jason Epstein has written, "the global village green will not be paradise. It will be undisciplined, polymorphous and polyglot." The academic world is and will undoubtedly remain highly pluralistic. Just within the domain of publishing, traditional journal production will remain with us for some time and the shift of those journals to electronic forms of dissemination is likely to continue.

Moreover, our institutions have a lot to gain economically in this transition from print to electronic publishing. Not only do electronic publications provide greater functionality for teaching and research than those in print, but also as a recent study published by CLIR has shown, there is good evidence that the considerable operational costs in libraries of ordering, receiving, processing, shelving, and circulating physical copies can be eliminated by a shift to electronic versions. In the aggregate across institutions, these potential savings may total in the tens of millions annually. It is worth noting, however, that these are the costs in the print world that represent our system of archiving. Cutting those costs without putting in place reliable archives that are committed to the academic mission remains a problem, and there is increasingly-and unfortunately-a widely held view among our academic institutions that savings are theirs to capture and reallocate, and that covering the costs of archiving digital assets for the long term is a responsibility for someone else.

In the face of this reasoning, let me cast the archiving problem in even starker terms. The shift to electronic publication in its current form represents a dramatic, jump-off-the-cliff shift in the academy from owning scholarly output to renting it. With subscription payments, institutions no longer buy journals, and the current form of licenses limit use so that more traditional, mostly regional initiatives for collaborative collection development and resource sharing across institutions are now next to impossible. Instead, the hundreds of thousands of dollars going out the door each year typically buy only a year's worth of access to the resource and only for members of a single institution. Not only have prices risen, but the material terms of the licensing deals are transforming the underlying infrastructure of resources available for teaching and research.

A growing number of senior officers of our colleges and universities-presidents, provosts, and chief financial officers-are beginning to question the huge risk to the future of their institution's core operations because of the growing dependence on a record of scholarship for which the institution is paying substantial sums but on which it has no real continuing claim. Library licenses often have a clause that reserves so-called "perpetual access" for the institution to the material to which it subscribes each year, but there are no reliable enforcement mechanisms in these clauses. Typically, publishers promise to transfer the material on a pile of CDs, but I am not aware that any such transfer has ever taken place, and it is unlikely that any institution has or will build the capacity to implement such a solution. Instead, collective action is needed, and Portico, a new organization will be emerging in the coming months as a trusted third party archive, offering a model for how such action might best be organized.

In fact, the Mellon Foundation, JSTOR, and, we expect, the Library of Congress are all making commitments to the development of Portico as a new digital archive that could help alleviate this substantial and growing business risk within the academy-and address the problem of providing enduring access to the scholarly record for the community as a whole. Libraries and publishers must both join in crafting the solution. Library supporters of the archive would require publishers to deposit journals in the archive as a matter of contract. Upon cancellation of a subscription by a participating library, the archive would provide perpetual access to that material to that library. After a period of time all material in the archive would be available electronically to all paying supporters of the archive. The business model would effectively turn a rent of journals into a buy for everyone participating, but the collective, collaborative action needed to implement this model requires that universities and colleges make it a priority to invest in the support of the archive at least a portion of the operational savings that they would accrue from the shift to electronic publications.

The collective action will likely also require mechanisms to audit and certify the archive to ensure that it is performing its archiving functions. Managing the institutional risks associated with continuing and growing investment in electronic resources is a looming problem that responsible financial officers, faculty, librarians, information technology specialists, and others in the academy must confront across the board, not just for electronic journals, and which they must together address collectively across traditional institutional and disciplinary boundaries. The Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP) has mounted a significant effort on this issue for a broad range of materials involving a large set of institutions, and you will hear more about this later in the workshop. But even for journals, a reliance on Open Access, however it is defined, is not an answer but just another form of the problem.

The Elephant in Our Midst and Its implications

An even more significant strategic issue that has the potential to profoundly and permanently disrupt the patterns of higher education is what Lorcan Dempsey of OCLC calls the "Amazoogle factor." It is now well known, and still deplored by some, that Amazon, Google, Yahoo, and other online systems are the first and sometimes only stops for students doing research. Faculty, too, have come to depend increasingly on these services. These organizations are now working closely with publishers to make the contents of current publications more accessible and "search

friendly." Google Scholar, which was announced last fall, may not be comprehensive in its coverage, but its ability to parse out citations from articles, among other remarkable features, shows how adept Google can be in addressing some of the more nuanced and specialized needs of scholars. And this is not the end. In mid-December, Google also announced that it has launched a massive retrospective digitization project based, at least initially, in five major research libraries.

One of the most common figures of speech that has appeared in public discussions of digitization over the last decade has been the invitation to imagine having the entire Library of Congress available electronically and accessible at the click of mouse. Google's investment in re-engineering the digitization process and of significantly reducing the costs so that it could undertake its own initiative means that the vision of digitizing the holdings of our largest research libraries is not only imaginable but may actually be within reach. The initiative and any competitive projects it might stimulate could be incredibly valuable for the public and for the academy in particular. But that Google is undertaking this effort, not for philanthropic purposes, but for business reasons, means that higher education-at least its library arm responsible for collecting, preserving, and providing access to content of scholarly significance-now has a formidable for-profit competitor with considerable resources and its sights set squarely on at least one core aspect of the higher education business. The deals that Google has made with its research library partners are just now beginning to be scrutinized in public, and those deals may not pass muster, as Microsoft's Corbis deals with museums fifteen years ago did not. But Google is different from Corbis, and its interest in library content resonates deeply with the interests of libraries in digitizing book content and enhancing search of those digital assets by more sophisticated means than catalog searching.

Let us assume that one way or another, massive digitization takes place. Among the big strategic questions for higher education would be how digital assets for scholarly communications should be organized in such an environment. These questions have scarcely been identified, much less aired and fully discussed. I am going to leave a number of these issues to one side tonight and instead highlight several other broad implications of Google's potentially disruptive influences on the academy and particularly on the ways that it manages its digital scholarly assets.

The "processed" publication. First, I want to draw attention to an idea that Joseph Esposito highlighted a few years ago in a *First Monday* article. For scholars, massive digitization and open access are not ends unto themselves. The central issue is whether scholars can advance knowledge in ways that were not previously possible. Scholars need to make use of digitized and open access materials. Esposito's insight is that at the highest level of generality, what unites our interest in digitization and open access in a digital world is that the material becomes "processable." That is, it is subject to computational processing: it can be indexed, manipulated, mined, aggregated, decomposed, built up, and so on by algorithm, and it is this "processability" that makes digitized objects and open access materials valuable to scholars.

Intellectual property. This brings me to a second point about intellectual property. The temptation is to throw up one's hands in despair at the massive cost of meticulously clearing the rights of every rights holder in an object to be digitized, and either to abandon digitization of copyrighted material altogether, or to engage in efforts-also costly but often not accounted for-to

stay under the radar of the copyright police. Google seems prepared to take the risk of violating copyright by displaying snippets of copyrighted material in search results and then handing off the searcher to the publisher or to a library where the searcher can obtain the full text legitimately in print or digital form. This approach would represent one of a growing set of initiatives, including Mellon-funded initiatives such as JSTOR, ARTstor, CIAO, ACLS's History-E project, the BiblioVault project at the University of Chicago, the Electronic Enlightenment at Oxford University, and New World Records, as well as others such as the ECCO project based at the University of Michigan and CLIR, all of which demonstrate that communities of users and publishers can find ways to create the needed trust and goodwill and agree to overcome the costly barriers of copyright to create highly useful digitized collections of research and educational materials.

Open access materials might lower costs to some who want to use them, but they will never comprise the full range of materials for scholarly purposes, and to that extent open access simply will not be a necessary condition of advancing scholarship. On the other hand, what do appear to be necessary to the future of scholarship are "processable" materials. There may well be an opportunity here to recalibrate licenses, rights, and even copyright law itself with a richer taxonomy of uses, many of which may currently be regarded by default as "piratical." Uses that support machine indexing, for example, may actually need to be redefined as legitimate, provided other protections are in place, because they have become the core infrastructure today for serving the U. S. constitutional principle of promoting "the progress of science and the useful arts."

Search. Third, I would highlight the need for new and expanded search and research capabilities. Google's indexing of full text would be generated by optical character recognition (OCR) and could greatly expand and facilitate basic searching and retrieval. Serious thought now needs to be given about ways that Google and other search engines could be used to achieve the metasearch and other service objectives we are trying to achieve, sometimes at great expense, in the catalogs of our local systems. However, we also need to be thinking beyond the local system catalogs.

The sheer volume of digitized material, for example, is going to require implementation of much more sophisticated indexing, searching, and filtering techniques, including broad application of computational linguistic and related statistical techniques as well as sophisticated techniques for filtering based on markup and thesauri, which would relate results to discipline-based concepts and concerns. Above all, there will be growing demand for mechanisms to link search results flexibly across systems in ways that resemble but will be fundamentally different from metasearching across catalogs. To provide a simple example: how easily could one search for related materials in ARTstor, and JSTOR, and, say, Readex Newsbank? Google or Yahoo may be able to respond to the basic demand for cross searching, but as scholars become more sophisticated in their use of these technologies, their needs will become correspondingly more specialized and discipline-specific in ways that it will likely be unprofitable to address for commercial companies aimed at the mass market. Search and information retrieval is a growth industry not only in the general economy but also for scholarly communications. Solutions that the large search engines cannot supply will have to come from search applications developed

within and for the academy, and finding these solutions should be a high priority for the academy to address.

Research methods. The fourth strategic area that I would highlight for you is the advance of new discipline-based research methods. The development of search technologies will drive the scholarly use of massively digitized resources, but scholarly use will also shape and guide the development of particular technologies and applications for specific disciplinary pursuits. Disciplines will need to develop new and specialized methodologies-an informatics of standards and practices-to identify, mark up, and explore the large volumes of digital information with which they each need to work: economists with tabular data in government publications; literature scholars with literary texts from various genres; social historians with contemporary accounts of various aspects of social life; ethicists with case studies of ethical dilemmas; art historians with evidence about the context of artists and their art; and so on. As scholars in various fields of study develop experience with these materials, the disciplines and sub-disciplines will need to develop and codify practice.

Over the next 3-5 years, if scholars begin to formulate how the use of these newly digitized materials could advance knowledge in their fields and begin to set discipline-based standards for how these materials should be organized for systematic use, then we will likely need to pave the way for three further types of intensive scholarly activity. Editorial activity will shift, field-by-field, to the markup and online annotation of digitized source materials to shape them for scholarly activity in particular disciplines. Tools will be needed to operate on these materials in discipline-appropriate ways. And given appropriately edited and marked up resources, and proficiency in new methodological techniques, scholars will begin to generate and report results based on research using these methods. These reports will refer systematically to digitized sources and may incorporate them in various ways. Researchers at the Institute for Advanced Technology in the Humanities at the University of Virginia, including Ed Ayers and Will Thomas, and elsewhere have been modeling new forms of scholarly practice like these. The results of this early work demonstrate that there will be a growing need for training and other forms of support from librarians and IT specialists in all forms of institutions as these practices take shape in discipline-appropriate ways and spread throughout the academy.

New collection emphases. The fifth strategic area for managing digital assets that I would bring to your attention is the need for dramatic shifts in the emphases in collection building and processing. If large quantities of published materials are available online through some common interface, it will be increasingly hard to distinguish libraries based on their holdings of these materials. Of course, scholars will always need access to the original artifacts for various purposes, and holding libraries will need to streamline the ways that they collectively manage these artifacts in offsite, and perhaps shared, regional repositories. Instead, libraries and their institutions will increasingly be distinguished by the special collections of rare and unique materials which they hold and by the scholarly services they provide for these materials. Special collections are often inaccessible or underprocessed, and the forms of description do not integrate well with other kinds of catalogs. Several institutions have been working together in recent years to develop innovative methods of appraising special collections for processing; others to simplify the cataloging. Building on these initiatives, support will be needed over the next 3-5 years for the development of more efficient forms of processing and description, and

revised standards so that special collections finding aids integrate more effectively in larger asset management systems.

Perhaps even more important is the need for more aggressive development of collections in new media. Recent and contemporary culture is documented in audio recordings, in still and moving images, and in various exclusively digital formats, such as geographic information systems, simulations, Web pages, and Weblogs. Scholars will increasingly need access to these materials for teaching and research. Concerted action is especially needed among libraries to ensure that these materials are actively and comprehensively collected and processed for scholarly use. Economies of scale, and the complexities associated with intellectual property rights management may prove-as it is proving for ARTstor and art images and New World Records and musical recordings-that individual libraries need more centralized, collaborative mechanisms to achieve these objectives.

Interaction between digital library and learning management systems. The last strategic issue that I would highlight is the need for more seamless interaction between digital library and learning management systems. There is a pedagogical trend to incorporate the use of primary sources and research methods more deeply in the curriculum of higher education, and this trend will likely continue, but will also vary by discipline. As scholars in different fields gain experience with and develop discipline-based methodologies for using massively digitized content, as well special collection and new media collections, they will need to incorporate the material and train students in the research methods. Demand will grow for deepening connections between digital library systems used for managing digital assets in various forms and combinations of licensed, digitized, and open access materials and learning management systems such as Sakai. Conversely, at least some of the content specifically created for teaching and learning will need to flow to digital library systems for long-term management and preservation. Essential for the effective management of the flows of content among digital library systems and between digital library systems are mechanisms, like Shibboleth, for building and expressing levels of trust between owners and users of the digital assets.

Conclusion

There is a view that the promise – or curse – of Google's activities is that they will make the management of scholarly digital assets within the academy largely irrelevant. I hope you can tell from the strategic issues that I have highlighted for you this evening that I find such a view to be spectacularly uninformed and shortsighted. Rather, the promise – or the curse -is that managing digital assets for scholarly purposes has become a vastly more interesting enterprise than it has ever been. It is increasingly possible for scholars to have unprecedented access to the resources they need to engage issues that have remained elusive or even unthinkable. The custodial challenge for us is to be both extraordinarily innovative and conservative at the same time: Innovative in that we must organize ourselves to take absolute best advantage of the opportunities; and conservative in that we must protect our gains and not screw up the scholarly process.

I promised to conclude by focusing on roles and division of labor. I cannot be prescriptive about who should do what. There is just too much to do. The need is great for imagination and

expertise to be applied wherever it can be found, from the largest institutions to the smallest. But allow me the following rough distinctions.

For many of the issues that I have highlighted, such as the shift in methodological practices in scholarly disciplines and the interaction of library-based content with pedagogical practice, much of the energy and support is inevitably local. Indeed, it will require levels of flexible and responsive service relationships between the faculty and library and information technology specialists, which the smaller colleges and universities have proven themselves to be extraordinarily adept at providing, compared to the larger research universities. In order to make and preserve the gains through local support efforts, I especially want to mention that generating funding is critical and requires that we make urgent, common cause with our faculty colleagues to produce a more disciplined, rigorous, and articulate public face than we have had in a long time about the value of humanistic study in a hostile and brutal world.

Having said that there is much room for local effort, it is worth pointing out that many activities, including initiatives involving mass digitization such as the Google activities, and even JSTOR and ARTstor, need to be organized centrally, largely for reasons of scale and economy. Activities involving uniquely held, special collection materials must take a more distributed approach and involve the holding institutions, but that does not mean that every institution must itself invest in digitization labs. Similarly, large-scale software development projects, like DSpace, Fedora, and Sakai, greatly benefit from central control to produce code efficiently and on-budget, and also require input and support from a broad range of users and experts, but each has generated slightly different mechanisms for opening its processes and engaging the broader community.

Beyond the need for local initiative and expertise, there is also a need for collaboration and collective organization involving shared financing and responsive governance at levels that are probably unprecedented. If massive digitization occurs, it would be imprudent for the institutions that provide the source material to cede complete control of the digitized versions to a commercial organization, and it is my understanding that each of the initial partners has ensured that Google will provide them a copy of the digitized versions. However, the volume of material to be digitized would be so large that, even with rapidly falling storage costs, no single institution would likely to be able to afford to store the digitized versions plus provide necessary backups. Moreover, no single institution would be the source of material to be digitized and the digital copies collectively would be of great use to the broad educational community. A collectively financed mechanism or organization-a “BookStor”-would probably need to be founded for the purpose of providing long-term preservation of the digital copies in trust for the community, if not actually for serving up digital copies to users once they are found and requested. Will Google agree to allow a “Bookstor” to come into being and will institutions rise up to support it?

The need for such collective organization raises another, larger question about how the academy can reorganize itself to accommodate efficiently and responsibly within its embrace entities that essentially outsource library and related functions that once were held closely within individual institutions. The California Digital Library is one model of outsourcing within a state system. JSTOR, ARTstor, and NITLE represent yet other models, and the Mellon and Hewlett

Foundations are experimenting with yet another in their jointly funded creation called Ithaka, which is designed to stitch together with common services ARTstor, JSTOR, NITLE, and a family of other scholarly support entities. These resources simply cannot take shape if they are imagined to be "one off," or ad hoc organizations. Presidents, provosts, deans, scholars, librarians, and technologists together must find ways within the larger academic community for their institutions to work together to realize the extraordinary economies of scale that are possible, and foundations like Mellon should not be seen as the "deep pockets" to which they turn to cover the costs of these entities, but as catalysts in the necessary effort to establish them financially and organizationally as new modes of ongoing operation in higher education.

Let me stop here and leave you with a cautionary tale from Adrian John's *Nature of the Book*. In nineteenth-century England, there arose a group called the Society for the Diffusion of Useful Knowledge. Worried that an educated working class could be a dangerous force in society, it resolved to swamp the country with cheap magazines-the Penny Cyclopedia and the Penny Magazine-that contained absolutely nothing "to excite the passions." To achieve this mission, the Society was the first group to make full, industrial use of the steam press, a remarkably cost-effective technology at large scale. By 1832, The Society's magazine was "by far the most extensively circulated periodical works that issue from the press." It estimated its readership at the then unprecedented figure of one million.

However, for all the attention to cost-effectiveness, critics of all persuasion attacked the project and it eventually failed. Conservatives were convinced that the project dispersed unnecessary ideas that might still prove dangerous. Radicals, on the other hand, complained that the magazines contained no really useful knowledge. Instead, they said, rather than meeting demand, the society sought nothing more than to "stuff our mouths with Kangaroos."

As we continue the discussions in this workshop and in the future about how most effectively to manage scholarly digital assets, let us not fall into the trap of the Society for the Diffusion of Useful Knowledge and lose sight of the ultimate objective: meeting demand for useful knowledge. Let us be on the lookout for the "Kangaroos."

Thank you very much for your attention.