UNIVERSITY ARCHIVES AND INSTITUTIONAL REPOSITORIES: CASE STUDY OF A WORK IN PROGRESS

Elisabeth Kaplan

I want to start by saying thank you to the National Archives, to the Associacao, to the organizers of the Congress, and especially to my good friend Lucia Maria Velloso de Oliveira. This is my second trip to Rio de Janeiro and I am so grateful for this opportunity to visit so many excellent institutions and meet so Brazilian colleagues.

I was asked to speak about how an academic organization has gone about planning for implementation of a program to manage digital resources. I am not a technology expert by any means, and I confess I have felt reassured by some themes we've heard throughout the week. As the presentations about programs at the National Archives in both the UK and the US demonstrated amply, building a digital program is not all about the technology -- that factors such as organizational context and institutional culture, partnerships, resources, and, I would add, the needs of archives users are just as important.

Let me begin by providing a little organizational context. The University of Minnesota is the second largest university in the United States, with a student body of over 50,000. That means that any discussion of digital preservation has to be considered in the context of this enormous scope and scale. The University is a public, land grant university which means that we exist to serve the citizens of Minnesota. And Minnesota has very clear legislation on public records, which means that with few exceptions, documents created by a public institution such as the University must be made publicly available. The University is governed by the Board of Regents, and according to Regents' policy, we must collect and preserve the records of the institution "regardless of format."

The University Archives has the usual roles and responsibilities: we serve as the university's institutional memory, providing stewardship of the historical record. We select, describe, preserve and provide access to that record. We provide support for research and teaching through these activities. We provide consultation on best practices for recordkeeping. And we support institutional planning, policy development, and decision-making.

In the digital age, these fundamental responsibilities don't go away; they just become larger and more complex. Fulfilling those means that we must change, and we must make difficult choices. This reminds me of Theo Thomassen's discussion of the evolution of the archival paradigm. The new paradigm develops in response to new questions that the old one cannot resolve. The new paradigm builds on, but does not replace the old one; it merely places it in a new interpretive framework.

And when I say "the digital age," I am referring to the current technological climate, in which first, the majority of information is born digital, and, second, our users expect online access to records in traditional formats. An example of each of those developments: one core resource that is now born-digital is the University's payroll. Every year for the past many decades, the University has produced a thick document listing every person ever employed at the University, with name, title, campus address, classification, and salary. This is tremendously useful information for a variety of reasons, and it is public information. But this information is now created digitally and stored in the University data warehouse, where it is not publicly accessible and not reliably captured for the long term.

An example of changing user expectations revolves around the minutes of the Board of Regents. This is the governing body of the University. The documentation produced at the Regents meetings is the first place to search for policy information, precedent to current policy. These are no longer distributed in paper format, and, while the minutes for the past few years are available on the Regents web page, they are not searchable, there is no access to the previous 150 years, and there is no plan in place for permanent access and preservation.

So the University Archives faces a mandate to develop the capacity to manage typical university archives documentation in digital formats, and, to deliver improved access to that core content.

To meet that need, we have launched a program called the "University Digital Conservancy" (UDC). It is in the planning stages now, which means we are developing policy, identifying early partners, selecting priority content, and working with a test instance of the software we have selected.

But our approach to developing the Digital Conservancy has been in many ways typically archival – we recognize the need to understand as much about our institutional context as possible, identifying core institutional functions and institutional priorities, potential partners and stakeholders. We have conducted focus groups with specific communities of stakeholders to learn

about their information needs. And we present the UDC not in abstract terms about loss of institutional memory and history, but in terms of creating valuable resources that will support the day to day work of our stakeholders. And to some degree, we let these records creators and users drive our decision making – rather than basing our decisions on what we as archivists feel is theoretically important. We do this because we know that the choices we make today will have to be supported in the institutional environment of tomorrow, and whatever we do, it has to be sustainable.

Of course our users' needs are diverse, and the institution is diverse, and the records and the formats are diverse, and there is no way to address all of the possibilities at once. While we would like to capture and preserve digital content regardless of format – email, blogs, web pages, wikis, we know that we have to start with a focus, and hope for some early successes that will demonstrate our value, gain support for our program so we can do more in the future (and here I am comforted by Adrian Brown's advice to "think big, start small"). We have to make some difficult choices about where to invest our resources. And, given that digital preservation is generally fraught with risks and unknowns, we decided to begin our UDC program with what in the digital libraries world is now a "tried and true" approach, that of the institutional repository. It's an application that has been used for many purposes over the past several years, but not yet for what we would consider traditional archival content.

I know that many of you are familiar with institutional repositories, or I.R.s., but I suspect that they are new to some of you as well, because from the start, the IR was almost exclusively the domain of librarians, scholars, and technologists. I've long thought, however, that the IR approach would work well with institutional archival content, and our program at Minnesota will be the first large scale IR implementation designed to meet archival needs. And so it may be useful to provide some background.

The IR movement was prompted by the convergence of several factors in the 1990s. A trend toward consolidation in academic publishing put scholars at a comparative disadvantage in terms of controlling rights to their own research output. The open access movement arose out of this concern, and the concept of a method for sharing research output directly with colleagues was attractive. The original conception was that of a technological infrastructure into which researchers would contribute content directly and from which content could be shared community-developed policies. The content would be preserved using large scale, stable, long

term storage, providing permanent handles through stable web addresses, and providing broad access through searchable, browseable collections of scholarly content.

The IR movement began as early as 1991, at Los Alamos National Laboratory, when physicist Paul Ginsparg developed the concept of a cooperative, online archive of physics preprints, which would allow physicists from around the world to share and comment on one another's work in advance of publication. The result was arXiv.org, now hosted by Cornell University. Arxiv has expanded tremendously (it now contains over 350,000 articles). http://arxiv.org/

Ten years later, DSpace, a collaboration between MIT and Hewlett Packard, became operational. And since then other IR platforms have been developed and implemented. IR's have been launched all over the world to provide long-term storage and persistent access to institutional and discipline-based research output. For example, here are a few IR implementations in Brazil:

Reposcom@PORTCOM - Communication's Sciences Repositories Portal, Brazil
http://reposcom.portcom.intercom.org.br/

Superior Tribunal de Justiça / Brazil
http://bdjur.stj.gov.br/dspace/

Universidade Federal do Paraná, Brasil
http://dspace.c3sl.ufpr.br/dspace/index.jsp

The best overview of IRs that I know of was written by Clifford Lynch, director of the Coalition for Networked Information, and published by the Association of Research Libraries in 2003.

"In the fall of 2002, something extraordinary occurred in the continuing networked information revolution, shifting the dynamic among individually driven innovation, institutional progress, and the evolution of disciplinary scholarly practices. The development of institutional repositories emerged as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content, and also scaling up beyond ad-hoc alliances, partnerships, and support arrangements with a few select faculty pioneers exploring more transformative new uses of the digital medium."

Beyond that, he continues:

"In my view, a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution. While operational responsibility for these services may reasonably be situated in different organizational units at different universities, an effective institutional repository of necessity represents a collaboration among librarians, information technologists, archives and records mangers, faculty, and university administrators and policymakers. At any given point in time, an institutional repository will be supported by a set of information technologies, but a key part of the services that comprise an institutional repository is the management of technological changes, and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services. An institutional repository is not simply a fixed set of software and hardware." http://www.arl.org/newsltr/226/ir.html

What I like about this description is that it underscores the concepts discussed earlier -- digital preservation entails a set of technological solutions, but must also occur in an organizational context, that explicit institutional commitment is required, and that the development of enduring cross-disciplinary partnerships and collaborations is crucial.

While DSpace was developed in response to interest from researchers, it was built on the expectation the faculty of a particular institution would be eager to share research output within an institutional framework, that the organizing principle would be institutional, and as more and more faculty members were prompted to share their work, a critical mass of rich MIT research material would be formed.

This did not turn out to be the case. Faculty members, it turns out, identify with a community of colleagues in their own discipline, not with colleagues from their own institution. A physicist at MIT wants to share her writing and research with a physicist at Berkeley, not with the historian in the next building over. Faculty members are evaluated by their peers, more so than their institutional colleagues. And so forth. There were other problems – the copyright questions remained thorny, publishers were not always cooperative, and uploading one's own content could be tedious. So contributions of content were not as plentiful as anticipated. (MIT has since worked to address those problems in a variety of ways.) The point, though, is that, once again, the problem wasn't with the technology, it was an unexpected reality stemming from organizational culture and human behavior. Content creators and content consumers did not behave as anticipated.

Perhaps not surprisingly, then, the most successful "institutional repositories" are those that are not institutional at all – rather they are built around communities of interest. "IR" became something of a misnomer, and we now talk about IR's and SR's, or subject repositories, or simply digital repositories, organized primarily around disciplinary lines. (A good example is the SR for agricultural and applied economics, http://agecon.lib.umn.edu/)

Archivists were for the most part uninterested in the development of institutional repositories, and only now beginning to explore the potential of the digital repository as a way to address the electronic records issues that have concerned the archival community since the 1980s. IRs were not developed for archives, but may prove to be a good fit for managing and providing access to our collections.

Imagine an IR of typical digital or digitized university content. The University, of course, owns its own information assets. We don't have the intellectual property and copyright concerns that arise with faculty research and publications, and we do not need to negotiate with publishers. The archives is charged with preserving and providing access to this content, so where faculty members might be unconvinced about the value of contributing to an IR -- that's our job. There are no tensions between institutional and disciplinary belonging – we exist only in the context of the institution. The IR is organized around the principle of provenance, where the records creator (e.g., the Office of the President) is a community, with sub communities, (Office of the Vice President for Academic Affairs), and so forth -- mirroring institutional organization and hierarchy. A mock up of a fully populated IR looks like one big finding aid, with multiple series and sub series.

The IR will not solve all of our problems or address all of our users' needs. Far from it. The software platform that we have selected, DSpace, is not a solution for managing complex objects such as web pages, or databases, or learning objects. It won't provide any enhanced functionality to those types of digital content. But it's a start, and a flexible and extensible one, and there is significant community support among institutions that have implemented DSpace.

In conclusion, there are some aspects to our program that are familiar, traditional archival responsibilities and concerns – selection, recruiting content, describing and providing access to information resources. But we are also taking an active role in crafting and developing this program, in partnership with our stakeholders. That collaboration will give us the ability to go

forward and do more. That is a model for evolution that promises to provide a tangible benefit to our community – the University – and a sustainable role for our archives. Together, those comprise the social and institutional context for our definition of a new archival paradigm.

It's risky and scary to embark on a project so fraught with uncertainty. We are experiencing exactly what Theo Thomassen described on the first day of the congress – living every day and coping with the very circumstances that will push against our disciplinary boundaries and force our professional paradigm to change. This is a kind of discomfort that we need to get used to, because if we know one thing, it's that the world of documentation will grow increasingly complicated, at a faster rate than we are experiencing now. This realization doesn't make the coping any easier, but I do believe that pushing the boundaries of our profession by finding methods for managing uncertainty is good for archives. It is a sign of a healthy discipline.