

# Towards a University of Michigan Institutional Repository: A Study and Prototype Proposal

Jim Ottaviani  
Cory Snavely

12 November 2003

## Executive Summary

This proposal, for a prototype Institutional Repository (IR), offers the University Library a means to enhance the U-M information environment and create a system to communicate and preserve the intellectual output of our scholars in ways not currently supported by more traditional library and publication models. Many universities and libraries have begun to investigate the repository idea, and their approaches cover a wide spectrum of services and content. A review of these projects and in depth interviews with Library selectors has resulted in our selecting the following as the main features of the IR we propose:

- a format-blind area for creating new scholarly works;
- a centralized resource that will allow U-M to represent the depth and breadth of its research and teaching to the local community, state, and world;
- a means of preserving the work of U-M researchers in a comprehensive way;
- a means of preserving scholarly work occurring at U-M that has never been captured, and;
- a resource that, when a particular community's area of the U-M repository is used in association with other subject-based repositories, will offer broad access to the output of specific scholarly disciplines.
- All of these features align with our current mission as a Library, and build on existing strengths in both traditional areas of collection development/management and leadership in digital library content and services. Specifically, the goals of the proposed Institutional Repository would include:
  - making a strong commitment to preservation and migration of data, with a well-defined, though modest in scope, commitment from and to all communities to preserve metadata about a particular IR item;
  - providing a service level that balances the need for communities to add content themselves with the need for format and metadata standards that allow us to meet our commitment to data preservation and migration;
  - creating a federated environment, which balances the convenience of peer-to-peer communication having the commonalities and strengths of a centralized service with standardized interfaces and service expectations;
  - offering an adjunct to, rather than a replacement of, the formal publication system. As such, the IR would seek shared rather than exclusive ownership of intellectual property;

- presenting creators with a spectrum of access options, with the ability for communities to constrain full access to their IR contents at a number of levels of openness, and;
- maintaining a strong preference towards content that has a high level of uniqueness to U-M and relationship to U-M campus activities.

We have identified potential partners who would provide content for the prototype. These partners, from across U-M and from small groups to Rackham as a whole, can provide us with the full range of formats, policy decisions, interface questions, and access issues we need to move from a prototype system to a full-service environment suited to the needs of the University as a whole. We have also acquired the necessary hardware (server) and software (DSpace) for the prototype. With these in place, we stand ready to begin creating this prototype when we are able to commit the personnel needed--1.75 - 3.0 FTE estimated for a two year prototype phase--to move this ahead.

## Introduction

The University Library, by proposing to create a prototype Institutional Repository (IR), seeks to enhance the current information environment and create a means to communicate and preserve the intellectual output of U-M scholars in ways that are not currently supported by more traditional library and publication models.

U-M has codified the need for offering an IR to both the U-M community and to the scholarly community at large by stating that "creating, communicating, preserving ... knowledge, art, and academic values" is central to our mission. Though an IR does not yet exist here in name, we have done an excellent job of creation, communication, and preservation of our own work. Creation and communication, sometimes on an *ad hoc* basis, occurs at the individual and departmental level. Communication (passively) plus preservation (actively) is the University Library's primary mission.<sup>1</sup>

As a result our scholars, departments, and libraries are world-renowned. To maintain this position, however, they rely on a complex publishing system dominated by a model where U-M scholars do their research, create new knowledge, and then typically give that knowledge away (or even, by way of page charges, pay to have it published) to commercial publishers. The loop closes when universities, including U-M, buy back this knowledge via their libraries in the form of books, conference papers, journal articles, videotapes, sound recordings, etc. This model has served us well for most of our existence, and the reward system for faculty relies on it.

As pressures on budgets become larger, though, and as direct, robust (and typically more informal) communication between individuals and institutions becomes easier, weaknesses in this model have become apparent. Universities can no longer afford to purchase all the material their scholars might need--or even buy back all the research done locally. Nor can we be sure we'll capture all of the interesting work that happens here in the first place. After all, many of the most

---

<sup>1</sup> The Library of course does not, in fact, limit itself to preserving and offering just U-M works, but performs these functions for scholars around the world.

vital aspects of the teaching, learning, and research community we foster at U-M--such as seminars, colloquia, and presentations by distinguished visitors--never get preserved.

So while we acknowledge and celebrate the diversity of research and experiences that make U-M what it is, we do not yet have a means for representing these things either to ourselves or to those outside our local community. The most natural thing to do for those who seek scholarship produced at U-M is to come here for it. Unfortunately, as noted above, in many (if not most) cases U-M does not retain the rights to--or sometimes even a physical copy of--works produced here. And if a copy is available on campus it may only be in a departmental filing cabinet, on a faculty member's computer, or stored somewhere on a network without the identifying information (metadata) needed to allow others to find and use it. The same is true of virtually every academic institution, so U-M scholars can't be sure that work done by a specific colleague elsewhere can be found at any given place either. As a result, the complex publishing system described above is supplemented by a complex system of interlibrary loans and unsystematic, informal networks of communication.

## **Institutional Repositories**

Fortunately, the means to start reasserting (co-)ownership and control of our own work have also become available in recent years, and there is a broad-based movement underway to create institutional repositories (IRs) to do this.

The IR concept, at its core, is simple: Take responsibility for U-M's intellectual output, both in terms of preserving it and communicating it to the scholarly community at large.

Many initiatives that address this need are currently underway. Locally, the Comprehensive Collaborative Framework (CHEF, including CourseTools/CourseToolsNG a.k.a. CTNG) WorkTools, and the Digital Asset Management System (DAMS) initiatives seek to facilitate and capture courses and projects at U-M. [CHEF, 2003; WorkTools, 2003; Hilton, 2003]. National organizations like the Scholarly Publishing and Academic Resources Coalition (SPARC), the Open Archives Initiative (OAI), and the National Science Foundation (NSF) seek to address this problem on an inter-institutional basis [SPARC, 2002; OAI 2003; Atkins, 2003].

U-M is an active participant in all, and the center--as in the case of CHEF and WorkTools--for some of these initiatives. However, as an institution we don't yet have the piece that can organize and preserve work present in systems like CHEF, DAMS, WorkTools, and work in areas not directly addressed by those projects. We also need to communicate that work to our peers and the world at large in keeping with our mission and supporting the visions that OAI, SPARC, and the NSF have put forth.

The proposed IR provides this missing piece. By creating an environment in which we can gather and federate the output of ongoing campus initiatives we can serve the needs of individual researchers, research groups, and the university as a whole. Acknowledging that the current publication system has many strengths worth preserving, U-M's IR would *not* seek to replace the

peer-reviewed, scholarly journal model for publication (a departure from SPARC's vision). Nor would it only supplement this model via sharing existing--and mostly copyright-free--content through open systems (per OAI). Instead, U-M's IR will work in conjunction with initiatives such as DAMS, CTNG and OAI to provide

- a format-blind area for creating new scholarly works that wouldn't have been possible previously;
- a centralized resource that will allow U-M to represent the depth and breadth of its research and teaching to the local community, state, and world;
- a resource that, when a particular community's area of the U-M repository is used in association with other subject-based repositories, will offer broad access to the output of specific scholarly disciplines;
- a means of preserving the work of U-M researchers in a comprehensive way, providing an alternative (or, at the very least, an addition) to the current scholarly publication model, and;
- a means of preserving scholarly work occurring at U-M that has, to date, never been captured.

To expand upon that last point, in addition to published works, U-M's creative output in the form of panel discussions, symposia, and selected student projects (to name a few), all of which are vital aspects of the U-M experience and few of which get preserved in a systematic or comprehensive way, will find a home that didn't exist before.

## Preparing for an Institutional Repository

### Technical aspects

Any number of hardware/software combinations can serve as a platform for an IR. DAMS, DLXS, OAI and even the library catalog can--and in the latter three cases, do--provide U-M creators with a place to put things and users a means to get them.

Software packages specifically designed to facilitate the creation of an IR is available, though. EPrints and DSpace are the two most commonly used, and show the most promise.<sup>2</sup> EPrints is free (GNU Project) software developed at the University of Southampton which many use to facilitate their IR projects. [See Appendix I for institutions using EPrints and DSpace.]

We propose using the DSpace software as the basis for the interface to our IR. DSpace is a digital asset/content management system developed by MIT and Hewlett Packard which manages and distributes digital items and allows for the creation, indexing, and searching of metadata associated with those digital items. It is designed to support the long-term preservation of the material stored in an IR. DSpace is also designed to make submission easy: DSpace

---

<sup>2</sup> Digitool from Ex Libris - <http://www.exlibrisgroup.com/category/DigiToolOverview> - may also be worth investigating in the future, though it does not have the relatively large user base that EPrints and DSpace currently do.

"communities" (e.g., U-M departments, labs, and research centers) can adapt the system to meet their needs and manage the submission process themselves if they wish. DSpace accepts all manner of digital formats, including documents (e.g. articles, preprints, working papers, technical reports, conference papers), books, theses, data sets, computer programs, simulations and other models, multimedia publications, etc. The DSpace system is freely available as open-source software. [DSpace, 2003]

The library has purchased and installed the necessary hardware to support a prototype IR, and is developing a timeline for installing the DSpace software on this system. The hardware is configured to support a modest pilot IR project, with a repository of approximately 0.5 TB in size.

### **Non-technical aspects**

With these pieces in place, the prototype will provide a test bed where we can demonstrate a proof-of-concept system for presenting the variety of scholarship that occurs here to local colleagues, peer institutions, and the public at large. But to put these pieces in place we need to address the human/political aspects of such a project as well as the technical. In fact, the technical aspects are trivial in comparison, especially given the Library's strengths in digital library infrastructure.

Starting with the ARL-SPARC-CNI "Institutional Repositories: A Workshop on Creating an Infrastructure For Faculty-Library Partnerships" held October 18, 2002, and continuing on through discussions with all U-M library selectors, we have identified four key factors for creating a successful IR. Note that three of the four have almost nothing to do with technology and everything to do with people.

The first and most important success factor is buy-in from producers. As mentioned above, scholars have a great deal invested in the publishing and production environment as it exists today. Peer-reviewed work offered via scholarly monographs and journals is the key to tenure, and supporting--or at least not subverting--this process is the key to short-term success for an IR. SPARC's long-term goal of breaking the hold of the costly peer-reviewed journal on scholarly publication is admirable, and we see an IR as one means of achieving it. But not yet. In the meantime, we propose working on a limited scale for the prototype, using a mixture of projects suggested by selectors. [As per interviews conducted in 2002-2003, summarized in Appendix II.]

The second success factor is a policy for inclusion in and access to IR contents. Such a policy would address

- types of work to include, addressing format, scope,
- rights and responsibilities of authors, communities, the IR (University Library) and U-M regarding preparation, submission, maintenance, and distribution of IR contents
- related issues including ownership, withdrawal/permanence of works in the IR, etc.

and would establish expectations for the IR as a whole. This policy would be the baseline from which individual communities would amend, modify, and interpret per their specific needs.<sup>3</sup>

The third factor is a strong commitment to traditional library values, specifically access to information, support of special collections, and preservation. We are recognized on campus and throughout the scholarly community as having this in abundance.

Finally, the fourth success factor is a robust digital library infrastructure. Though this was characterized as being the one heavily reliant on technology, it too has a strong human component. We are fortunate to have both the equipment and the expertise here in the Library to make an IR possible today, though we note that both are at or near capacity already.

## **Building a Prototype Institutional Repository**

### **Existing models and where U-M is today**

As described above, a number of institutions have begun to create IRs. Many are well beyond the conceptual stage and into the prototyping and implementation process. Appendix I offers a selected list of the most active institutions, and gives an idea of what they are making available.

The institutions listed are beyond the conceptual stage in their implementation, and many are operating--though as the numbers indicate, they operate with a limited scope. All are going ahead without trying to solve all the problems and create all the policies in advance--this is a sensible approach as we enter what is new, yet familiar territory given our experience with IR-like projects.

At this point, U-M has the server hardware available and installed, and Library staff have been contacted, interviewed, and have enthusiasm for the IR. See the summary in Appendix II for the thoughts and recommendations of Library selectors. As a result of these interviews, some policies/positions have been identified, again per discussions with selectors and a review of other IR policies.<sup>4</sup> We recommend:

---

<sup>3</sup> Two examples: 1) Our policy might state that only "finished, publication-ready" works are appropriate for the IR, as this was a common requirement suggested by selectors during the interviews--though just as common was a wide range of definitions of what "finished" and "publication ready" mean. Academic disciplines would and should interpret this requirement differently. 2) Different communities may have different needs when it comes to making works intended for publication in other venues also available through the IR. So, escrow policies along the lines of "only metadata available to all/U-M campus/local IR community for the first X years after submission" will use different values for X.

<sup>4</sup> Given our choice of DSpace for the software component of our IR, the DSpace policy at <http://libraries.mit.edu/dspace-mit/mit/policies/> is an appropriate starting point. Also note the

- A strong commitment to preservation and migration of data, with a well-defined, though modest in scope, commitment from and to all communities to preserve metadata about a particular IR item.
- A moderate service level, balancing the need for communities to be able to add content themselves without requiring Library involvement from an item's inception (but rather a hand-off of responsibility whenever possible) with the need for format and metadata standards that will allow us to meet our preservation and migration commitment.
- A federated environment, which balances the convenience of peer-to-peer communication and the commonalities and strengths of a centralized service with standardized interfaces and service expectations.
- An adjunct to, rather than a replacement of, the formal publication system. As such, the IR would seek shared rather than exclusive ownership of intellectual property.
- A spectrum of access options, with the ability of communities to constrain full access to their IR contents to personal, group/community, U-M-campus, or worldwide.
- Flexibility in terms of what a community defines as a finished (refined/vetted) work
- A strong preference towards contents that have a high level of uniqueness to U-M and relationship to U-M campus activities.

The devil is of course in the details, but with these general principles in mind we are confident that we can create a baseline policy that will meet both the specific needs of communities and U-M's overall desire to create, communicate, and preserve our knowledge, art, and academic values. A baseline policy that would follow from these ideas could then be added onto by individual communities as needs arise, in consultation with our selectors and staff.

## **IR prototype participants**

Based on these general policy considerations, we have identified the following interested participants for an initial prototype:

- Performing Arts Technology (Professor Mary Simoni, School of Music and Chair of the Department of Performing Arts and Technology): The School of Music, under Prof. Simoni, would like to create a digital archive of faculty and graduate student performances recorded at the School. In addition to Professor Simoni, the School of Music also has a faculty member who has responsibilities as a recording engineer who would be an important participant in the project.
- The Engineering Research Center for Reconfigurable Manufacturing Systems (Professor Yoram Koren, Director): The Engineering Research Center (ERC) group is currently creating a local collection of papers published by its researchers. The IR prototype will use the work already done as a basis for exploring how to offer published works in a useful, rights-sensitive, way.

---

many DSpace participants appear to be using MIT's policies as their basis as well. (See MIT's "About DSpace@MIT" - <http://libguides.mit.edu/c.php?g=176372&p=1158986> - for their current policies.)

- Kresge Business School (Tomalee Doan, eLibrary Director): The Business School has begun to place videos/multimedia on its website, and serve it through their video streaming library at <http://www.bus.umich.edu/Technology/>. Working with Kresge, we will investigate associating robust metadata and serving bandwidth-intensive content through the IR, working with libraries outside of the main U-M system, and working with DAMS, which also plans to experiment with the Business School's videos.
- Transportation Research Institute (Bob Sweet, Head of UMTRI Library): The University of Michigan Transportation Research Institute (UMTRI) and the Library have been digitizing all the UMTRI papers, and are currently making them available as a stand-alone database on the web. These papers are of high interest, and comprise a large and growing collection. We will use this collection to investigate migration/incorporation of an existing collection whose digital aspect was developed by the Library.
- Rackham Dissertations (Rex Patterson, Asst. to the Dean for Information and Technology Services): Making U-M dissertations (both Ph.D. and Masters) available online has long been of great interest to a broad range of people, both on campus and off. The timing appears to be right to move this forward.<sup>5</sup>

## Next steps

To move forward, we need to prepare the server for the DSpace software, install it, and begin to use it. We also need to commit to deploying staff (our existing infrastructure) to do the additional outreach and detail work to provide actual content for the IR.

The hardware is in place, and the DSpace software is free. For this prototype phase, which we foresee as lasting two years, we will need a configuration and design team comprised of

- a leader (one person, 0.5-1.0 FTE);
- liaisons to the communities/participants (two people, 0.75-1.0 FTE), and;
- technical staff to do configuration, and interface work (one person, 0.5-0.75 FTE).

The leader and one of the liaison roles could be filled by the same person working full-time, thus realizing some efficiency.

In addition, promotion and publicity, both in terms of graphic design and liaison to internal and external library partners, are important tasks as well. They may indeed be essential--this would add another person at 0.25-0.5 FTE, depending on whether we chose the low or high estimates for the above leader, and what that leader's skills might be.

A specialist in digital conversion and preservation, at 0.5-1.0 FTE would also be ideal, so as to better address preservation and migration issues from the outset. Indeed, given the expertise,

---

<sup>5</sup> Recent investigations by Matt Stoeffler of OpenOffice (<http://xml.openoffice.org> and <http://www.xml.com/lpt/a/2001/02/07/openoffice.html>) indicate that new tools exist that promise making conversion from Microsoft Word to more stable, portable, and robust formats easier than they have in the past.



investment, and strong architectural components we already have in digital library operations such as preservation and migration--particularly DLPS and SPO--there are clearly opportunities for integration with an IR effort, and addressing them would be a priority for the pilot process.

Finally, there will certainly be a draw on our existing infrastructure. In addition to the advisory roles that Sr. Managers, PSC, etc. will play, groups like NISC will provide a critical and active role in providing direction and guidance to the project leader.

Though rough estimates, we believe these to be at or near the minimum we should devote to this project, given its potential for affecting the campus information environment as a whole. Again, the Library is in a unique position: We have all the talent, skills, and technology infrastructure required to take a leadership role in creating an Institutional Repository. If we make this a priority, our experience and status in the digital library community makes us an excellent candidate to help lead the national library community as a whole in the direction of asserting greater ownership of our campus' intellectual output.

## **Conclusion**

The University and the University Library are in a strong position to move forward on an Institutional Repository. Our core values speak to the preservation and access needs of our scholars. These values have been with us since the Library's inception, and we have demonstrated an ongoing commitment to them throughout our transition to digital production and delivery. And when it comes to the online world, our strengths in digital production and dissemination via MOA, SPO, OAI, and DLXS (among others) are also known worldwide.

The time is right to bring this considerable expertise and experience to bear on local content, especially those publications or events that are not likely to be captured anywhere else. The Digital Asset Management System (DAMS) initiative on campus can address some of this, but the Library is in a unique position to complement the DAMS idea by offering long-term and public access to those resources that should be preserved and presented to a wider community of scholars. DAMS also indicates the high interest in U-M information assets, at the highest level. The current project is, relatively speaking, narrowly focused, though. Through preliminary discussions with library subject specialists and selected groups on campus, we have broadened the definition of "assets," and have identified key partners for moving ahead with an Institutional Repository prototype.

These partners, from arts and technology and from small groups to Rackham as a whole, can provide us with the full range of formats, policy decisions, interface questions, and access issues we need. When we partner them with a relatively modest, dedicated team from the Library, we will create a prototype system and position the Library to make that system into a full-service environment suited to meet the needs of the University as a whole.

## References

- [Atkins, 2003] Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, Daniel E. Atkins, Chair, January 2003, [http://www.communitytechnology.org/nsf\\_ci\\_report](http://www.communitytechnology.org/nsf_ci_report) (accessed 15 April 2003).
- [CHEF, 2003] CHEF Information Site, <http://www.chefproject.org> (accessed 15 April 2003).
- [DSpace, 2003] FAQ: DSpace: MIT Libraries, [dspace.org/what/faq.html](http://dspace.org/what/faq.html) (accessed 17 April 2003).
- [Hilton, 2003] James Hilton, "Digital Asset Management Systems," *EDUCAUSE Review*, vol. 38, no.2, March/April 2003, 52-53.
- [Lawrence, 2001] Steven Lawrence, "Articles freely available online are more highly cited," *Nature*, vol. 411, no. 6837, p. 521, 2001.
- [OAI, 2003] Open Archives Initiative, <http://www.openarchives.org> (accessed 15 April 2003).
- [SPARC, 2002] "The Case for Institutional Repositories: A SPARC Position Paper," prepared by Raym Crow, 2002, available at [http://www.sparc.arl.org/sites/default/files/media\\_files/instrepo.pdf](http://www.sparc.arl.org/sites/default/files/media_files/instrepo.pdf) (accessed 6 February 2014).
- [WorkTools, 2003] U-M WorkTools (accessed 15 April 2003).

## Appendixes

### Appendix I: Institutional Repositories

The following is a selected list of institutions with significant (defined as having more than 200 items/documents accessible) IR efforts underway.

#### [EPrints.org](#)

- ANU: Australian National University (Australia)  
<http://digitalcollections.anu.edu.au/>
- Caltech (USA)  
<http://coda.caltech.edu>
- CCSD: these-EN-ligne (France)  
<http://tel.archives-ouvertes.fr/index.php?langue=fr&halsid=hch0aqmpjjna80tbds72hcro40>
- IUBio: An archive of biology data and software maintained at Indiana University Biology department (USA)  
<http://iubio.bio.indiana.edu:7780/>

- LU:research from Lund University (Sweden)  
<http://eprints.lub.lu.se/>
- UNITN-eprints: University of Trento - Italy  
<http://eprints.biblio.unitn.it/>
- University of Pittsburgh (USA)
- Archive of European Integration, PhilSci Archive  
<http://aei.pitt.edu/>  
<http://philsci-archive.pitt.edu/>
- University of Queensland ePrint Archive (Australia)  
<http://eprint.uq.edu.au/>
- University of Southampton: Department of Electronics and Computer Science (U.K.)  
<http://eprints.ecs.soton.ac.uk/>

## **DSpace**

The DSpace Federation includes the following universities:

- Columbia University (planning)
- Cornell University  
<http://dspace.library.cornell.edu/> : live, 1 paper available
- Massachusetts Institute of Technology  
<https://dspace.mit.edu/> live and robust, though only one collection (Sloan) getting frequent additions)
- Ohio State University (planning)
- University of Rochester (planning)
- University of Toronto  
<https://tspace.library.utoronto.ca/> called TSpace: live and "robust" with just over 200 items
- University of Washington (planning)

## **Others**

- Blekinge Institute of Technology (Sweden)  
<http://www.bth.se/eng>
- California Digital Library (USA): System software: custom with Berkeley Electronic Press (Bepress)  
<http://repositories.cdlib.org/escholarship/>
- CERN Scientific Information Service (Switzerland)  
<http://cds.cern.ch/>
- Lulea Institute of Technology (Sweden)  
<http://www.ltu.se/?l=en>
- Universität Dortmund (Germany)

- System software: Hyperwave ( <http://www.hyperwave.com/e/> )  
<http://eldorado.uni-dortmund.de:8080/>
- Universität Konstanz (Germany) OPUS System  
<http://www.ub.uni-konstanz.de/kops/>
- Universität Stuttgart (Germany): OPUS System  
<http://elib.uni-stuttgart.de/opus/>
- Utrecht University (The Netherlands)  
<http://www.uu.nl/EN/library/igitur/Pages/default.aspx>
- Virginia Tech, Digital Library and Archives (USA)  
<http://scholar.lib.vt.edu/>

## **Appendix II: Summary of interviews with University Library Selectors (Conducted Fall-Winter 2002-2003)**

### **What should an Institutional Repository contain?**

#### **Presentations, colloquia, seminars, performances, recitals and events sponsored by and/or held at U-M**

Transcripts, video, audio, or even simply abstracts from the many and varied events sponsored by U-M departments. Examples include the Raoul Wallenberg lectures, Charles & Roy Eames lectures in architecture, the Penny Stamps lectures in Art & Design, "Grand Round" presentations at the Medical School (which are overviews of the field by experts), Bishop lectures at the Law School, UMS and Power center works (or records of same) "Dialogs in Diversity", "Saturday Morning Physics", and other programs sponsored by the various centers on campus.

#### **Theses and dissertations**

Capturing masters and senior honors theses comprehensively, and presenting those theses along with our PhDs is an important way to present U-M scholarship to researchers, at peer institutions, alumni, and prospective students as well.

#### **Nominated/best-quality student work**

(Related to the above.) From Hopwood winners to honors theses to musical compositions to engineering design projects nominated by faculty members, undergraduates (as well as graduates) produce high quality work which deserves preservation and dissemination. Again, this will be of value to both researchers and students, who will see U-M as providing a resume-building opportunity via this presentation.

#### **Working papers, formal publications (at the very least, listings of same)**

From working/white papers and studies (e.g. sponsored by ISR and frequently quoted in mainstream media) through to items that are produced by U-M scholars (either visiting or permanent) and formally published elsewhere, a record of these works at minimum, and full-text/image as the ideal will help us present what U-M is about. CVs of faculty can provide lists

#### **Raw materials associated with formal publications/finished products**

(Related to the above.) Raw data, original lab notebooks, surveys, questionnaires, interviews/transcripts, art works in progress, etc. All in support of the above to enrich scholarship and allow for others to build upon work that's already been done.

#### **Publications based at U-M**

Examples include the *Philosopher's Imprint*, the new journal being launched by The Center for Afro-American and African Studies, *Dimensions* from the College of Architecture and Urban planning. and no doubt many others. (Even the *Gargoyle*!)

#### **Out-of-copyright materials**

Brittle books, papyri, and other selections from our collections that contribute to the public good. Out-of-print materials from the U-M press would also be worthwhile to add. (DSpace is doing this for MIT press books.)

#### **Exhibitions**

The Botanical Gardens, Libraries, Museums, and others mount exhibitions where a great deal of scholarly work is captured--in the form of e.g. descriptions--but only for a brief time. Rotating these on an opening page of the IR would provide an eye-catching display and incentive for return visits, perhaps.

#### **Departmental histories**

Tracing the intellectual evolution and interests of faculty through departmental histories (as well as collection histories, perhaps) is something difficult to do currently. This area has a strong overlap with Bentley's interests (monthly and annual reports for departments, agendas and minutes of committee meetings, syllabi, announcements of lectures and public events, newsletters, etc.)

#### **Course materials**

Archiving syllabi, websites, reading lists, interactive tutorials, and other items created and/or stored within e.g. CourseTools (cf. MIT's Open Knowledge Initiative, which as a public institution U-M arguably has a greater obligation than they to make open)

#### **Other**

Websites (similar to histories, provide snapshots of the U-M at a given time). grant proposals, information about U-M (press clippings, articles, University in the news)

The common theme throughout the discussions was that the focus should be on U-M scholarship. In other words, if it was done here (regardless of whether the producer is a full-time, permanent faculty, staff, or student member of the community) it should be represented in the IR. Unlike the traditional Library role, we would not buy items for the repository, but rather gather things already owned and produced here, doing conversion if needed to make the material more widely accessible.

With the above material at its core, an overall view of the campus work would be possible for the first time. Having such a view could strengthen intercampus collegiality, help make the case for funding at the state legislative level, and heighten U-M's profile on the national and international scene.

#### **How unique should items in the IR be?**

The question of what type of material should go into the repository comes with a corollary: What authors/creators of material should be included in the U-M IR. The vast majority (though it wasn't unanimous) of people interviewed think that only materials unique to U-M should be represented. In other words, if something was produced here, or by someone associated with U-M at the time of production, it belongs in the IR.

A possible exception would be for special collections we've built, which become U-M creations in and of themselves by virtue of their unique composition.

#### **What formats should we support?**

- Print
- ASCII
- PDF
- html
- XML/marked-up/mark-up-able
- some paper, incl. monographs
- Images
- slides--35mm, microscope

- possibly some copy photography using archival standards on e.g. dpi, bit depth
- Audio
- students will want MP3, not just archival
- Video
- MPEG and un-compressed
- with transcriptions would be ideal
- dance with multi-camera shoots
- some animation needed to show use of tools
- conversion from 16mm
- Course Materials
- PowerPoint
- authorware,
- Director
- CourseTools
- other websites
- Data
- raw ASCII, which most data archives require
- databases
- spectral and structural chemical, for which competing standards exist
- uncompiled computer code
- GIS and spatial
- 3-D objects
- architectural models

### **What quality (level of finish) should we require?**

Inherent to the discussion of working papers and supporting materials above (indeed, inherent to all the possible areas for inclusion) is the idea that the IR will contain scholarly work that may not be in its complete and final form. There is in fact great value to making working papers, raw data, transcripts, and other supporting materials available to others who wish to build upon the work done by U-M researchers.

(Note that it is easy to argue for extremes in this so as to leave nothing at all for the IR to do: Unfinished work and raw materials do not belong because it doesn't represent the most developed thoughts of the researcher, nor are these things easy to digest for an outsiders... Published papers and books do not belong in the IR because they are easy--or at least relatively easy--to obtain from other sources.)

Every scholarly community--indeed, every individual--will have different standards and needs for sharing pre-publication work. Some will only want to present work they would consider ready for formal publication, others may want to present a complete trail/history of the research.

A model policy as proposed by the IR administrators could be offered as a starting point, but with the above in mind, I suggest creating policies for including materials in the IR on a community-by-community basis with the direct input from members of those communities.

### **How permanent should this be (once in the IR, should documents stay forever)?**

A clear consensus here also: For a repository to be worthy of the name, it should make a commitment to permanence. While there are cases where an item might have to be removed for e.g. plagiarism, academic dishonesty, or legal reasons, we expect such cases to be rare--and in those cases where the actual item is no longer in the IR it must leave a clear "footprint" (via metadata and a note regarding its withdrawal). All of the above harks back to one of print's most compelling features: the ability to cite and return to material with confidence that it will be there for the next use and the next scholar.

That said, clear mechanisms for being able to supersede existing items is essential to making the IR a living entity of maximum use to researchers.

### **Should we restrict access to the IR?**

While the preference would be not to have to restrict access in any way, the nature of an IR is such that we foresee a variety of access levels ranging from personal (for items that may be created in the context of the IR but aren't yet ready for publication) through to completely open to the world. A mine/ours (i.e. my group's)/U-M's/everyone's set of rights will require rights management mechanisms (including authentication/authorization and perhaps even payment mechanisms) and the ability (via policy) to move items in the repository between various categories--ideally progressing towards fewer/no restrictions as these items age.

With this in mind for the actual content of the IR, the metadata associated with those items would have significantly fewer restrictions. In the ideal case it would be globally searchable even when the item(s) this metadata describes was not be globally accessible. This is a model we already use for materials we acquire for the libraries.

The following overlapping/complementary lines cover the complete range of access levels:

Mine--No metadata assigned, only accessible to the creator.

Ours--"Publication" occurs here. Content is not available to anyone, metadata available to U-M, group, and individual

U-M--Content is available to U-M community (and those visiting campus, in most cases). metadata accessible to all



Everyone--All metadata, all content associated with it, accessible to anyone. (Note that some items may not get/need metadata--lack of metadata need not preclude something being included in the repository).

### **What tools should we provide users of the IR to find/manipulate/use/create things there?**

A fundamental theme in terms of tools was not so much the manipulation of individual items (downloading data, reformatting text, panning and zooming on images) but of the view of the IR itself. Multiple ways of interacting with the IR are desirable. A U-M-wide view makes the most sense from an administrative perspective, and perhaps also from the perspective of a new user (one not familiar with U-M's breadth of research, or inexperienced in the area of scholarship they wish to pursue). However, for the experienced researcher, or for someone who is interested in working within a discipline related to one they're already associated with at e.g. another institution, the ability to look at only a particular section of the IR is essential. Building inter-university communities of researchers through IRs is a goal we can support if our IR is a federation of individual repositories, all operating around core capabilities (e.g. metadata) and interface, but able to be viewed separately as well.

Any tool implies we can take what's in the IR and re-use it, so rights management (including payment for e.g. viewing and printing) again becomes an issue. This ties into the DAMS initiatives already underway on campus.

A major theme throughout the conversations so far is that the IR should provide scholars with the ability to create new work within the IR environment and use the environment as a medium for scholarly discourse (e.g. comment on the work of others). This implies an obvious crossover with projects such as WorkTools and Chef as well as library-based projects such as OAI.

### **What issues do you see in terms of gathering/providing access to the contents?**

The first of the major issues identified are acquisition and ownership. The two are linked closely, in that we will need some means of acquiring material for the IR, and if the IR asserts ownership of the materials that is seen in any way to be exclusive, or to preclude our researchers from operating within the traditional reward system (peer review, publication, tenure) we won't succeed. So, we'll need strong, producer-centered policies at the university level to ensure contributions. Aspects of creating and implementing this policy will include educating U-M researchers on the rights they currently have (and often sign away to publishers), creating and meeting expectations for long-term delivery of IR contents, and creating expectations for contribution by U-M scholars. (Including students, who have a different relationship with U-M than our faculty and staff, and thus will have different expectations and needs.)

On a more technical note, the cost of storage, retention, and serving materials in the IR can be relatively modest (cf. peer-to-peer file sharing) and in fact is likely to decrease over time. However, to create a useful research tool we'll need to offer rich descriptive metadata for the IR's

contents. Though it doesn't necessarily have to be done at ingest, and may to some degree be assigned automatically or by authors, it is expensive work, and is the second major issue.

Policies for inclusion (the traditional selection role performed by librarians) and retention (the traditional weeding role of librarians) are joined by policies for migration of rights/access. All of these will need to be community-based--librarians have a role here, but not a role played in isolation from the actual producers. (Since unlike our catalogs and other resources that mainly represent purchased collections, the IR as described above will be a living, working environment.)

As a sidebar, the relationship with Bentley is an important one, and making sure our policies and architecture complement each other is important. (As a possible way to divide the potential material, perhaps if something is "born in paper" it will most likely be the Bentley's, and if it's "born digital" it may be the IR's.)

### **What collections/archives exist already that might provide a back-file?**

- CAAS has a large music collection, and the director is particularly keen on having it cataloged and made accessible. [Afeworki Paulos]
- Ghana oral tradition audio tapes (proposed, but does not yet exist here). [Afeworki Paulos]
- "Michigan Codex" (10th century Pentateuch) [Elliot Gertel]
- Individual, unique titles, archival collections. [Elliot Gertel]
- Rabbi Levin papers (1 box worth) has strong donor interest--consists of class notes, sermons (mostly in Yiddish, Zionist information. a particularly important letter from Rabbi Kook). [Elliot Gertel]
- Joseph and Marie Adler collection--very wide range, large (70 boxes worth). [Elliot Gertel]
- Moses Gomberg papers (currently some are in Bentley)--first Jewish instructor at Michigan, discoverer of free radicals (Elliot has strong personal interest, having written a paper on him). [Elliot Gertel]
- TBZ (Jewish fraternity) archives. [Elliot Gertel]
- Penny Stamps lecture videos [Annette Haines]
- Newspaper collections (uniquely held?) [Julie Herrada]
- American Committee for the Protection of the Foreign Born archives (120-150 boxes)--mostly print, unique, heavily used, stored off-site, and important. (and EAD finding aid exists) [Julie Herrada]
- ICPSR is doing some of this--they take data sets in a "public format" (personal identity information is stripped) and make it available. [Jennifer Nason Davis] [JoAnn Dionne]
- Medical data, "Tecumseh Study" [JoAnn Dionne]
- Historical information (e.g. Regents' proceedings), historical course catalogs, reports (e.g. Public Safety, campus architect, salary)--basically, things Bentley has [Marija Freeland]
- The Business School/Library is putting their working papers online. [Marija Freeland]
- CRLT (if we don't want to limit to U-M-centered work) has archives. [Marija Freeland]
- Institutional memory (in the form of e.g. oral history--this would be a new collection area made possible by IR AND MAY HAVE FUND-RAISING VALUE) [Marija Freeland]

- Large mass of data waiting, but much of the finished product needs to go to large, existing commercial or societal publishers. [David Peck]
- Lecture tapes (e.g. by Buckminster Fuller). [Rebecca Price]
- Print volumes exist of the history of health sciences faculty. [Lauseng, Martin, Redman, Shipman, Townsend]
- Medieval and renaissance manuscripts--600 volume collection, color is important, all we have is an old bibliography from the 1930s [Kathy Beam]
- Islamic manuscripts (in Arabic, Persian, Turkish)--800-900 [Kathy Beam]
- Dutch materials [Kathy Beam]
- Philipino materials, for which there is a current NEA grant. [Kathy Beam]
- The U-M website (though it's not archived) [Denise Schoene]
- Math historical collection [JoAnn Sears]
- Saturday Morning Physics videos exist for the last few years. [JoAnn Sears]
- SNRE has a Center for Sustainable Systems--wanted the library to catalog their masters' theses in particular, which deal mainly with Michigan environment. [Sara Rutter]
- SNRE's "Endangered Species Update" (SPO is digitizing) [Sara Rutter]
- Naval Architecture has some reports that might be candidates. [Leena Lalwani]
- Edison collection of sheet music (only 14% cataloged), high profile [Reynolds, McConnell]
- Stanley quartet (reel-to-reel) concert performances, currently underway as a two-phase project with the MediaU [Reynolds, McConnell]
- Composers forum concerts, spotty coverage exists [Reynolds, McConnell]
- Current papyri (20k records, ~15k image), ideal will be 50k full images when complete [Traianos Gagos]
- Psychology chair initially wanted to scan and archive (via CD) what's in peoples' file cabinets--don't know what (if anything) was done with this idea [Darlene Nichols]
- CAAS has plenty, and it's not just from faculty--student papers, videos from invited speakers. [Chuck Ransom]
- SNRE masters theses (some are paper-based, some are project-oriented) [Dottie Riemenschneider]
- Biostation papers. [Dottie Riemenschneider]
- OAlster-able things (e.g. from the Kelsey Museum) [Jennifer Nason Davis]
- SPO materials--including French encyclopedia Bryan is working on. [Bryan Skib]
- Anne Waldman videos (special collection on beat poetry) [Jeff Pearson]
- Depository videos from the state of Michigan [Jeff Pearson]
- Cafe Shapiro anthology [Forrester, Gaither, Kolekamp, MacKintosh, Martin, Tenofsky, TerHaar, Tuckett] [Peggy Daub]
- Ray Tanter's online courses [Grace York]
- Historical library documents [Grace York]
- Michigan Enslan [Grace York]
- Regents proceedings [Grace York]
- Enrollment figures [Grace York]
- Alumni records (may be touchy/difficult) [Grace York]
- Clippings files about U-M [Grace York]
- Islamic manuscript collection [Beau Case]

- Kelsey museum materials, including field reports from excavations, surveys, photos, artifacts [Beau Case]
- Ethnic studies materials (with aspects as per Kelsey materials above) [Beau Case]
- Latin inscriptions (we have largest collection in Western Hemisphere) [Beau Case]
- Center for Japanese Studies publications, e.g. quarterly newsletters [Kenji Niki]
- Medical departmental annual reports have comprehensive publications lists in them [Arndt, Rana, Schnitzer, Rosenzweig]
- Axelrod collection--copyrighted material in Public Health, also unpublished papers (15+ filing cabinets) [Lauseng, Jo Keller, Allee, Alana O'Neal, Faiks, Michael McLaughlin, Look, Pulsifer]
- Jonas Salk's laboratory slides, etc. [Lauseng, Jo Keller, Allee, Alana O'Neal, Faiks, Michael McLaughlin, Look, Pulsifer]
- "Findings", HBHE (published at/via U-M?), publications [Lauseng, Jo Keller, Allee, Alana O'Neal, Faiks, Michael McLaughlin, Look, Pulsifer]
- Myron Wegman documents [Lauseng, Jo Keller, Allee, Alana O'Neal, Faiks, Michael McLaughlin, Look, Pulsifer]
- Campus plans (architectural) for renovation of SPH buildings [Lauseng, Jo Keller, Allee, Alana O'Neal, Faiks, Michael McLaughlin, Look, Pulsifer] belong in Bentley?
- Davison Institute papers [Hsu, Wan]
- Oriental art archives [Hsu, Wan]
- Hussey papers [Hsu, Wan]
- Depository items (per our mandate, previously much stronger, to capture U-M materials). Science library has many shelves of this. [Patricia Yocum]
- Madeline Albright lectures [Doan and Sokkar]
- MacNally Lecture series [Doan and Sokkar]
- Working papers (already doing this using a DLPS tool) [Doan and Sokkar]
- Interdisciplinary Child Welfare Training Program [Karen Reiman-Sendi]
- Social Work Distinguished Lecture Series [Karen Reiman-Sendi]
- Global Program on Youth information [Karen Reiman-Sendi]
- Project for Research on Welfare, Work, and Domestic Violence [Karen Reiman-Sendi]
- Smart Girl website [Amy Robb]
- GARP (a longitudinal survey of schoolchildren in Prince George County) [Amy Robb]
- Hopwood winners [Tom Burnett]
- Ensians [Tom Burnett]
- Philosophers Imprint [Scott Dennis]
- CV information [Scott Dennis]
- masters theses [Scott Dennis]
- Olin papers on Law and Economics [Barb Garavaglia]
- Law is doing a lot of this already (with meeting notes, correspondence, draft laws, working papers, etc.) [Barb Garavaglia]
- Slice of Life (sharing instructional resources by Susan Stenass at Cornell) provides model? [Pat Anderson]
- Peter Sparling dance videos [Wallach, Bartlett]
- Solar car group [Wallach, Bartlett]
- Duderstadt collection [Wallach, Bartlett]

- UMS had pressing need right now, strong interest--wants to build a site w/legacy data, continually renewed information; they are very high profile (lots of alumni/public loyalty) [Wallach, Bartlett]
- Expand on the Community of Science information [Karen Downing]
- Economics working papers [Kathleen Folger]
- Anne Waldeman symposium materials (we get complete videos by/of her, and readings) could add on transcripts, parts of her archive [Kathleen Dow]
- Marge Piercy will be at U-M in 2004--similar interest/need for the Waldeman materials [Kathleen Dow]
- Proceedings of the Alternative Press Symposium [Kathleen Dow]
- Youtie papers (which match up with our papyrus collections) [Kathleen Dow]
- John Sinclair collection (at the Bentley) [Kathleen Dow]
- Zelma Weisfeld archives (theater professor who will offer material including fabric swatches, costumes, production bibles) [Kathleen Dow]