



3-26-2008

Opening Doors to Open Scholarship: A White Paper Exploring the Role of a Digital Object Repository to Support and Preserve the University of Kentucky's Digital Assets

Lisa Carter

Ohio State University - Main Campus, carter.1088@osu.edu

Mary H. Molinaro

University of Kentucky, molinaro@uky.edu

Rebecca Ryder

University of Kentucky, becky.ryder@uky.edu

Kathryn Lybarger

University of Kentucky, kathryn.lybarger@uky.edu

Beth Kraemer

University of Kentucky, kraemer@email.uky.edu

See next page for additional authors

Follow this and additional works at: https://uknowledge.uky.edu/libraries_reports

 **Click to open a feedback form in a new tab to let us know how this document benefits you.**

Part of the [Archival Science Commons](#)

Repository Citation

Carter, Lisa; Molinaro, Mary H.; Ryder, Rebecca; Lybarger, Kathryn; Kraemer, Beth; Helm, Brian; Thomson, Mary Beth; Soward, John; and Scaggs, Dierdre, "Opening Doors to Open Scholarship: A White Paper Exploring the Role of a Digital Object Repository to Support and Preserve the University of Kentucky's Digital Assets" (2008). *Library White Papers*. 1.

https://uknowledge.uky.edu/libraries_reports/1

This White Paper is brought to you for free and open access by the University of Kentucky Libraries at UKnowledge. It has been accepted for inclusion in Library White Papers by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

Authors

Lisa Carter, Mary H. Molinaro, Rebecca Ryder, Kathryn Lybarger, Beth Kraemer, Brian Helm, Mary Beth Thomson, John Soward, and Dierdre Scaggs

Opening Doors to Open Scholarship

A White Paper Exploring the Role of a Digital
Object Repository to Support and Preserve the
University of Kentucky's Digital Assets

Prepared by
Institutional Repository Work Group
University of Kentucky Libraries

November 29, 2007
Revised March 26, 2008

Table of Contents

PREFACE	3
EXECUTIVE SUMMARY	4
INTRODUCTION	5
DIGITAL LIBRARY/DIGITAL OBJECT REPOSITORY	6
DIGITAL PRESERVATION	8
INSTITUTIONAL REPOSITORY	11
SCHOLARLY COMMUNICATION AND AUTHOR RIGHTS MANAGEMENT	13
ELECTRONIC JOURNAL PUBLISHING	15
INFRASTRUCTURE AND RESOURCES NEEDED	17
SUMMARY OF RECOMMENDATIONS WITH SUGGESTED ACTION STEPS	18
CONCLUSION	20
BIBLIOGRAPHY	21
APPENDIX A: WORK GROUP CHARGE	23
APPENDIX B: AUDIT CHECKLISTS	24
EXAMPLE FROM SECTION A: ORGANIZATIONAL INFRASTRUCTURE.....	24
EXAMPLE FROM SECTION B: DIGITAL OBJECT MANAGEMENT	25
EXAMPLE FROM SECTION C: TECHNOLOGIES, TECHNICAL INFRASTRUCTURE & SECURITY	26
APPENDIX C: SAMPLE FACULTY SURVEY	28
APPENDIX D: OTHER RESOURCES CONSULTED	30
APPENDIX F: GLOSSARY	32

Preface

The Institutional Repository Work Group was charged and began working in early June 2007. The charge to the group may be found as Appendix A. The purpose of the group was to examine the issues surrounding the development of an institutional repository and what role the library would have in that development and management. The membership of the group was intentionally kept small in order for the group to be nimble and easily able to complete the task at hand. The group was initially chaired by two co-chairs, Lisa Carter and Mary Molinaro. Other members of the group were Rebecca Ryder and Kathryn Lybarger from Preservation and Digital Programs; Beth Kraemer and Brian Helm from Information Technology; Mary Beth Thomson, Associate Dean of Collections and Technical Services; and John Soward from University of Kentucky Information Technology. During the course of the work Lisa Carter left the University and was then replaced on the Work Group by Deirdre Scaggs, University Archivist.

The Work Group began by gathering and examining articles about and examples of institutional repositories. Members divided the sections of the charge to the Work Group and made presentations to the group followed by in-depth discussion of the topic. Each Work Group member then drafted sections of the report after having been informed by the discussion. The sections were then put together and edited by the entire group.

The paper is written to take the reader from a general overview of digital object repositories and issues of digital preservation through the more specific issues encountered in an institutional repository followed by issues related to specific types of content. There are recommendations in each section. There is a summary of recommendations at the end of the paper with some suggested action steps also outlined.

The appendices of the paper include the charge, the audit checklist, a sample survey, additional resources the work group consulted, and a glossary.

Executive Summary

University faculty, departments, and colleges are producing more digital content than at any time in the past. This content may live on servers across campuses, may be maintained on departmental websites or may be stashed on flash drives. Regardless of where the digital content resides, a serious risk exists that the content may disappear over time due to data corruption or failure to execute data migrations. As a result, academic libraries across the United States now recognize the need to collect, manage, and preserve the cultural, intellectual, and scholarly memory of their respective communities.

Scholarly research is extensive and widely varied at the University of Kentucky. Now, more than ever, the results of that research are being created, stored and disseminated in digital formats. Not only are scholars increasingly creating traditional materials such as books, journal articles, and theses and dissertations in digital formats such as PDF and Microsoft Word, but they are also creating digital audio/video materials, massive data sets, and interactive software.

The Institutional Repository Work Group examined the issues related to digital content creation in a research environment and concluded that the University will benefit from ongoing, sustained access through a digital object repository (DOR). The DOR will function as a system to include current, disparate digital object repositories and will create the opportunity for deposition and maintenance of new collections. The institutional repository, a subset of the DOR, will offer faculty, staff, and students a vehicle, methodology, and a “safe place” for their scholarly efforts. The overarching DOR will incorporate the IR, e-journal content, digitized resources from the Libraries’ collections and other digital resources created at the University. Therefore, shared support is appropriate, standards are required, and collaborative management is essential. The development, maintenance, and preservation of the DOR will require campus-wide administrative support with specific tasks and budgets assigned to both the University Libraries and UK Information Technology. This trusted digital repository will stimulate and encourage the creation of open scholarship and will ensure the persistence and usability of the University’s priceless digital assets.

The Work Group studied the issues and crafted seven recommendations in support of a digital object repository which will foster the development of an institutional repository for University of Kentucky scholarship. Recommendations are summarized on page 19.

Introduction

University faculty, departments, and colleges are producing more digital content than at any time in the past. This content may live on servers across campuses, may be maintained on departmental websites or may be stashed on flash drives. Regardless of where the digital content resides, a serious risk exists that the content may disappear over time due to data corruption or failure to execute data migrations. As a result, academic libraries across the United States now recognize the need to collect, manage, and preserve the cultural, intellectual, and scholarly memory of their respective communities.

Scholarly research is extensive and widely varied at the University of Kentucky. Now, more than ever, the results of that research are being created, stored and disseminated in digital formats. Not only are scholars increasingly creating traditional materials such as books, journal articles, and theses and dissertations in digital formats such as PDF and Microsoft Word, but they are also creating digital audio/video materials, massive data sets and interactive software.

This practice is not isolated to computer science and other classically technical disciplines. Musicians and composers use digital composition programs, some that are score production programs, but others that are “companions” in the music creation process. Education experts may produce learning modules with embedded grading criteria. Medical students may produce videotapes of patients with a given symptom to describe it in a way that would not be possible in words. These materials represent the scholarship of the University and should be archived in a digital object repository (DOR) and accessed through an institutional repository (IR).

A vast amount of digital material is also being created through digitization of library collections. UK Libraries’ Preservation and Digital Programs Department has created several terabytes of images from photo collections, newspaper images from microfilm, and oral histories from audio cassettes. These materials should be archived in a digital object repository. These materials, however, do not constitute University of Kentucky scholarship and therefore should be accessed through an interface other than an institutional repository. Currently, the Kentuckiana Digital Library serves as the repository for the UK Libraries’ Kentucky-focused digitized content.

Even the day-to-day administrative work of the University is increasingly accomplished in digital format. Official communication is done by e-mail and University records are created, stored, and accessed electronically. Minutes of University Committees are word processed and shared electronically, and websites for academic department come and go. While a small portion of permanent University records are sometimes placed in the University Archives, UK Libraries’ role in the management of these records is largely advisory. The Work Group acknowledges the importance of these digital materials and the issues surrounding them, but they are outside the scope of this discussion.

University-generated digital objects may be in a wide variety of formats, some open-source and some proprietary. For example, text documents may be in plain text, RTF,

LaTeX, PDF, or Microsoft Word; images may be JPEG, GIF, PGM, or TIFF; audio may be MP3 or WAV; and video may be AVI, MPEG4 or QuickTime. Software may be available as source code, as a binary program that runs on some computer system, or as specialized software (such as a learning module) that runs within a given framework. Additionally, new formats will emerge and need to be accommodated. The formats used for digital objects will have a great impact on the DOR's ability to migrate and preserve the information in them.

In this White Paper, the Work Group outlines the components and functions of a digital object repository, discusses the unique features of an institutional repository, examines the issues of scholarly communication, rights management and e-journal publishing, defines digital preservation management, and outlines the resources needed to establish a trustworthy and enduring digital object repository for the University of Kentucky.

Digital Library/Digital Object Repository

A digital library is a means to ingest, store and disseminate digital objects in a variety of formats and types. The Association of Research Libraries (ARL) has identified five elements common in most definitions of a digital library:

- The digital library is not a single entity;
- The digital library requires technology to link the resources of many;
- The linkages between the many digital libraries and information services are transparent to the end users;
- Universal access to digital libraries and information services is a goal;
- Digital library collections are not limited to document surrogates: they extend to digital artifacts that cannot be represented or distributed in printed formats.¹

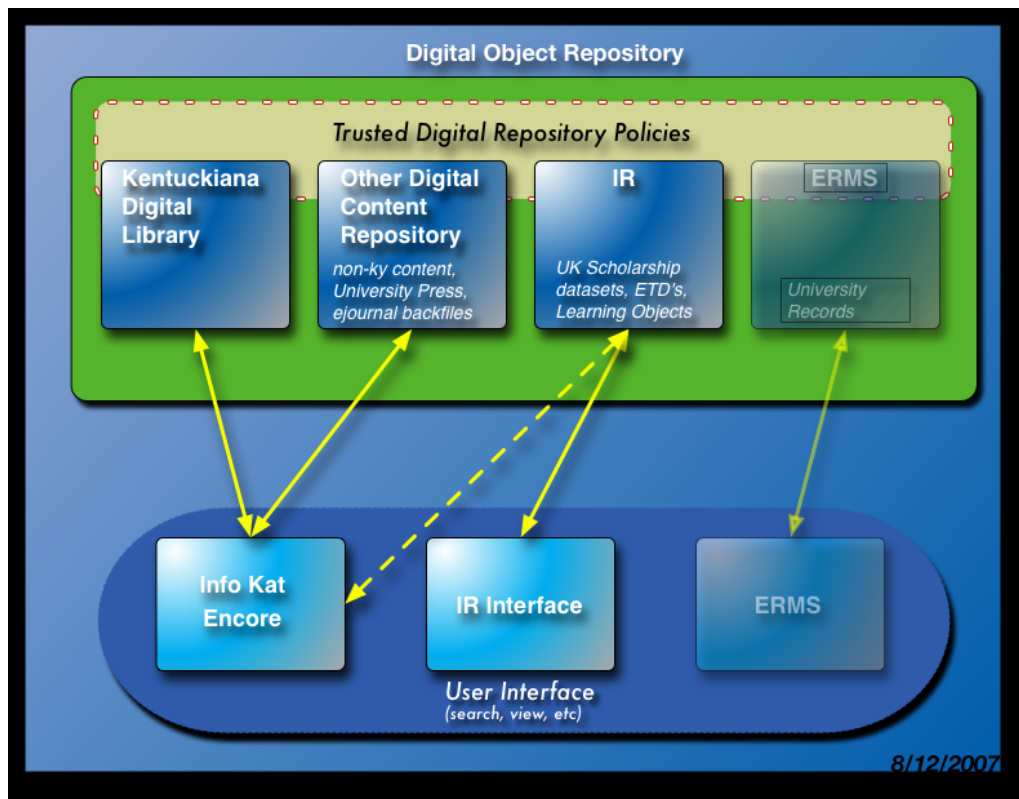
A digital library houses many types of collections. To represent this in the broadest terms and to best represent the needs of the University of Kentucky, the Work Group refers to this entity as a Digital Object Repository (DOR). Essentially, anything that can be represented electronically can be found in a DOR. To accomplish access to a DOR, multiple systems may be utilized, each custom-focused for specific collections or digital object types. These systems are usually linked by a common interface for seamless access to all digital objects.

The DOR must also be able to preserve the integrity of its collections. Without a means to do so, the DOR would not be sustainable. There is some flexibility in how this is accomplished. For instance, the entire DOR may conform to articulated and agreed-upon policies and standards for preservation, or each individual system within the DOR can have its own set of policies and standards. The key to success is that the standards are clearly stated, uniformly applied, and institutionally supported.

¹ Association of Research Libraries. "Appendix II," *Realizing Digital Libraries: Membership Meeting Proceedings*. Boston, MA: 1995. <http://www.arl.org/resources/pubs/mmproceedings/126mmappen2>

In addition, without a central repository for digital objects, content creators are left to store and distribute their data using whatever means available to them at the time of creation. This is problematic for several reasons. Without DOR management, data can become lost or corrupted with no ability to recover it. In fact, the data can be lost without anyone even realizing its disappearance. Without DOR management, we also lose the ability to track usage of the University’s digital content, and we lose the opportunity to track the “provenance” of the scholarship, i.e., how it changes over time. Without usage and provenance data, the disparate custodians of University scholarship lack a strong voice to recruit needed resources for central DOR management. Moreover, we miss out on opportunities to promote the University’s research and achievements. Finally, without a central repository, the University’s digital research and scholarship is difficult to find if not totally inaccessible.

A model of a DOR for the University of Kentucky is represented in the following diagram:



In this model, the UK DOR is represented as the common system that provides seamless access to all digital objects. The Kentuckiana Digital Library (KDL) houses published materials and primary source materials relating to the history and culture of Kentucky. The IR is the home of University scholarship. The Electronic Records Management System (ERMS) refers to the management of the University administrative records, which is outside the scope of this White Paper. The symbol identified in the diagram as Other Digital Content Repository refers to a yet un-built repository which could be the home for electronic journal archives, digitized special collections, publisher-created

content, and any other digital artifacts that the University needs to manage but do not fit within the scope of the KDL, IR, or ERMS.

Recommendation

Without a central repository for digital objects, content creators are left to store and distribute their data using whatever means available to them at the time of creation. Given the fragility of digital content and the value of these digital assets to UK, the Work Group recommends that a digital object repository be established with adequate funding and support as soon as possible. The DOR will include an IR for faculty, staff, and student scholarship and will absorb and manage the digitized content currently produced and stored disparately throughout the University.

Digital Preservation

For a digital object repository to fulfill its function and promise as a storage and access resource, it must have an articulated digital preservation plan. According to an American Library Association Work Group on Defining Digital Preservation²

Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.³

Digital preservation policies document an organization's commitment to preserve digital content for future use; specify file formats to be preserved and the level of preservation to be provided; and ensure compliance with standards and best practices for responsible stewardship of digital information.⁴

The ALA Work Group on Defining Digital Preservation concluded that digital preservation actions and strategies address content creation, content integrity and content

² At the Midwinter Meeting of the American Librarian Association, the Work Group on Defining Digital Preservation was tasked to develop a definition of digital preservation for use by the library and archival community. Currently, the draft document has been accepted by the ALA Digital Preservation Discussion Group and is under consideration by the Preservation and Reformatting Section (PARS) Executive Committee. Wider input will be solicited from the Association of Library Collections and Technical Services (ALCTS) sections before the final version from PARS moves forward. The draft can be found at the *Defining Digital Preservation* blog at <http://blogs.ala.org/digipres.php>

³ From the draft of the ALA/ALCTS/PARS document defining digital preservation. The quotation cites the "Medium Definition." The draft can be found at the *Defining Digital Preservation* blog at <http://blogs.ala.org/digipres.php>

⁴ From the draft of the ALA/ALCTS/PARS document defining digital preservation. The quotation cites the "Long Definition." The draft can be found at the *Defining Digital Preservation* blog at <http://blogs.ala.org/digipres.php>

maintenance, all attributes that characterize a trustworthy institutional repository.
Expanding on these criteria:

Content creation includes:

- Clear and complete technical specifications
- Production of reliable master files
- Sufficient descriptive, administrative and structural metadata to ensure future access
- Detailed quality control of processes

Content integrity includes:

- Documentation of all policies, strategies, and procedures
- Use of persistent identifiers
- Recorded provenance and change history for all objects
- Verification mechanisms
- Attention to security requirements
- Routine audits

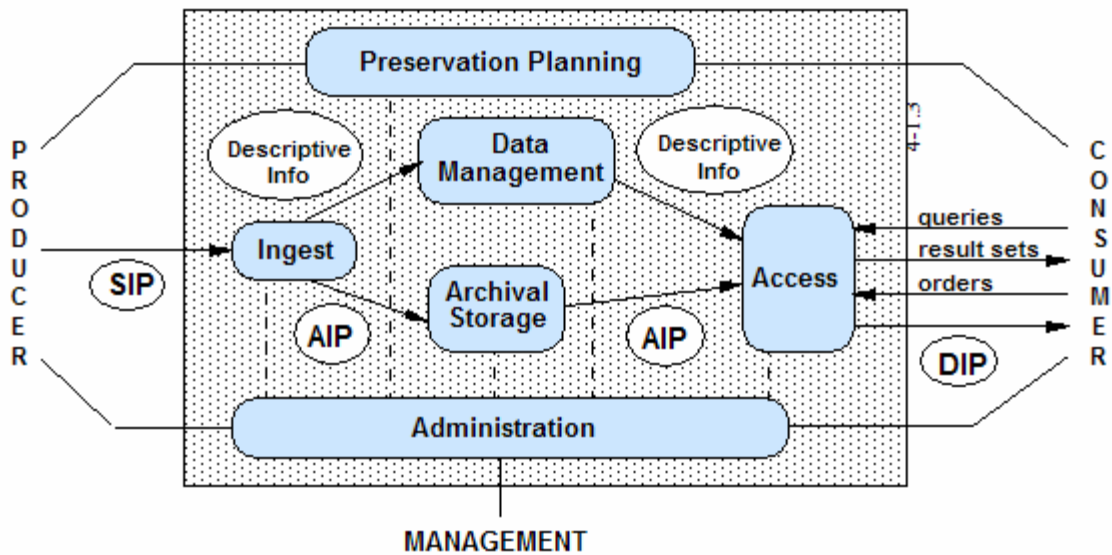
Content maintenance includes:

- A robust computing and networking infrastructure
- Storage and synchronization of files at multiple sites
- Continuous monitoring and management of files
- Programs for refreshing, migration and emulation
- Creation and testing of disaster prevention and recovery plans
- Periodic review and updating of policies and procedures⁵

The seminal research that underpins the definition is the Open Archival Information System (OAIS) reference model describing the functions, environment, and data characteristics for data preservation. OAIS refers to Archival Information Packages (AIPs), Submission Information Packages (SIPs) and Dissemination Information Packages (DIPs) which are all managed data packages with appropriate metadata and specifications that meet the preservation and access functions of the producer and consumer environment. The OAIS model represents the SIP, AIP, DIP functions as a forward- and backward-compatible, dynamic system framed within an active preservation planning administrative environment. At Cornell University's web tutorial on Digital Preservation, the OAIS Reference Model is represented in this diagram:⁶

⁵ From the draft of the ALA/ALCTS/PARS document defining digital preservation. The quotation cites the "Long Definition." The draft can be found at the *Defining Digital Preservation* blog at <http://blogs.ala.org/digipres.php>

⁶ *Digital Preservation Management*. (tutorial) Cornell University, 2003. <http://www.library.cornell.edu/iris/tutorial/dpm/foundation/oais/index.html>. The original OAIS Reference Model can be found at the *CCSDS Recommendation for an OAIS Reference Model*. Consultative Committee for Space Data Systems, Washington, D.C.: CCSDS, 2002. <http://public.ccsds.org/publications/archive/650x0b1.pdf>



The OAIS framework, researched and identified by the Consultative Committee on Space Data Systems of NASA, forms the intellectual infrastructure for the “Trusted Digital Repository” model created by the Research Library Group (RLG) and OCLC.⁷

RLG and OCLC took the theoretical OAIS reference framework and “translated” it into a model for digital object repository preservation. Their report, *Trusted Digital Repositories: Attributes and Responsibilities* (TDR), outlined six widely-encompassing features including:

- Compliance with OAIS model
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technical and procedural suitability
- System security
- Procedural accountability

The TDR Report explores and explicates each attribute in detail. Clearly, the underpinning of a trustworthy digital repository is strong administrative and financial commitment for ongoing support and growth of the repository.

Following the acceptance of the OAIS/TDR model for digital object repository preservation, RLG and the National Archives and Records Administration collaborated to create a tool to “certify” that digital repositories meet the TDR criteria.⁸ A trusted digital

⁷ *Trusted Digital Repositories: Attributes and Responsibilities: An RLG/OCLC Report*. Mountain View, CA: Research Libraries Group, 2002
<http://www.oclc.org/programs/ourwork/past/trustedrep/attributes01.pdf>

⁸ *Digital Repository Certification*. <http://www.oclc.org/programs/ourwork/past/repositorycert.htm>

repository cannot simply declare itself to be trustworthy. Trustworthiness has to be built into infrastructure of the repository, and it must be demonstrably certifiable according to a standardized and accepted audit system. This objective approach to certification builds the trust of the repository's creators, custodians, consumers, producers, and funders over time and documents "layers" of trustworthiness as the entire repository system grows and matures.

RLG and NARA vetted their report, *Audit Checklist for the Certification of Trusted Digital Repositories*, in 2005 with an international community of stakeholders. The Center for Research Libraries (CRL) and OCLC published the final version, *Trusted Repositories: Audit and Certification: Criteria and Checklist (TRAC)* in February 2007. The report...

...represents best current practice and thought about the organization and technical infrastructure required to be considered trustworthy and capable of certification. It establishes a baseline definition of a trustworthy digital repository and lays out the components that must be considered and evaluated.... It discusses the envisioned uses of this document, and the principles underlying the application of the criteria. Finally, it documents criteria that trustworthy repositories will be able to meet, providing explanations and examples.⁹

TRAC includes sections about organizational infrastructure, digital object management and technologies, technical infrastructure and security. The report discusses each section, explains the criteria in the checklist, and provides examples of evidences. The elements/criteria are formatted into a spreadsheet/checklist for gathering the evidence examined, the findings, the observations, and the overall result of meeting, or not meeting, a criterion. One characteristic of the audit checklist is that it features the regular scheduling of audit procedures as one of its criteria. Therefore, the audit process is an ongoing feedback mechanism that supports growth and demonstrates accountability of resource utilization. See Appendix B for sample pages of the TRAC audit checklist.

In reviewing the checklist vis-à-vis what we knew about UK's repository structure, the Work Group concluded that very few of the administrative and organizational criteria of TRAC were met. This finding points to the need for a centralized DOR empowered to organize, administer, set policy, and deploy resources to provide ongoing access to and preservation of DOR content.

Recommendation

Develop the DOR to comply with the OAIS reference model. Without robust and proactive preservation administration, the University's digital assets are rendered imperiled, if not totally unsustainable.

Institutional Repository

⁹*Trusted Repositories: Audit and Certification: Criteria and Checklist* <http://www.crl.edu/PDF/trac.pdf> pages 2-3.

Today, many universities are meeting the digital storage and access needs of their faculty and students by establishing institutional repositories. As defined by Clifford Lynch, “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.”¹⁰

Institutional repositories can take various forms within an academic institution, but they generally share the same goal—assuring the preservation of resources produced by the local community consistent with the mission of the university.

The IR is most successful if it is focused on a specific set of services that are institutionally defined. The success of the IR depends on how locally driven it is. For example, if the institution is known for the production of its faculty’s musical compositions, then music preservation would be a service upon which to focus. But the IR can also provide a long term commitment to manage, disseminate, and preserve many resources. Some of the most common digital objects are faculty scholarship including white papers, conference papers, pre-prints and post-prints. Other research may be in the form of data sets, digital art, and electronic theses and dissertations.

The first core feature of the IR is that its content is digital. The IR provides a centralized location to store, preserve, and make university scholarship accessible. It is community driven and focused; its success lies in allowing faculty to decide what goes into the repository—not librarians. The goal of the IR is to create a durable and permanent centralized archive which takes the responsibility off of the individual, department, or college to maintain the digital files that can quickly become obsolete. With that said, best practices and standards should be established and implemented at the beginning, and the IR should not be seen as a “digital dumping ground.” The IR is also widely accessible and should not be a dark archive where only our community members have access.

Another core feature of the IR is a submission mechanism for the content. Submission may be done by using a web-based form or another method of upload such as FTP. The metadata about content, structure, and the creative process may be submitted with the digital object. This process can also be tailored by the individual institution. Access control determines who can see the content of the IR as there may be some files that will have limited access versus open access. An IR allows for increased discovery because search engines such as Google will crawl the IR and make it easier for users to find the content. Having a centralized system built on standards specifically designed to support discovery through search engines distinguishes an IR from informal systems where digital content may be scattered across various campus web pages, personal computers, or servers. Finally, the centralization of digital information facilitates efficient and accountable file preservation. Most importantly, long-term digital preservation requires a long-term institutional commitment to preserve digital scholarship and to migrate it as technologies evolve.

¹⁰ Lynch, Clifford. “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital,” *ARL: A Bimonthly Report* no. 226 (February, 2003) <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>

Recommendation

The Work Group recommends that a survey should be conducted to identify the different formats actually used by current UK scholars. Information from the survey will inform the data managers of the research tools used at the University. Sample questions for such a survey are included in Appendix C.

Recommendation

Fundamental to the success and the life of the IR system is the active participation of the University's faculty through the contribution of content. Therefore active and strong collaboration among all stakeholders is essential. The Work Group recommends that marketing, outreach and education be established to "sell" the IR and to provide the tools and training necessary to make it successful.

Scholarly Communication and Author Rights Management

In the discussion about an institutional repository, the Work Group looked at the expansive issues of scholarly communication, rights management, and electronic journal publishing. The next two sections of this White Paper focus on those issues.

The Association of College and Research Libraries (ACRL) defines scholarly communication as "the system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use. The system includes both formal means of communication, such as publication in peer-reviewed journals, and informal channels such as electronic listservs."¹¹

There are two important aspects of scholarly communication that are directly impacted by an institutional repository: the dissemination of research data and digital content and the preservation of that scholarship.

An IR provides an avenue within the university for storage and promotion of scholarly communication. UK's mission includes the statement "Serves a global community by disseminating, sharing, and applying knowledge."¹² Universities are realizing the importance of promoting the scholarly contributions of faculty, staff, and students to numerous internal and external bodies including those that fund research. Recent open access discussions have included support by numerous universities of the Federal Research Public Access Act of 2006 (FRPAA). The UK University Senate voted to endorse FRPAA at their October 9, 2006 meeting. FRPAA addresses the need to maximize public access to federally funded research. Research, meeting specific criteria, would be deposited into a federal digital repository or a repository meeting particular conditions, thereby providing free public access, interoperability, and long-term

¹¹ See Scholarly Communications definition at:
<http://www.ala.org/ala/acrl/acrlpubs/whitepapers/principlesstrategies.cfm>

¹² See University of Kentucky 2006-09 Strategic Plan at:
<http://www.uky.edu/ucapp/plan.htm>

preservation. Many funding agencies have begun to allow and encourage open access to published research within certain parameters. IRs are being used as one method of improving the dissemination of and access to scholarly output.

On January 11, 2008, the National Institutes of Health (NIH) announced a new NIH Public Access Policy to ensure that the public has access to the published results of NIH funded research. This policy that becomes effective April 7, 2008 requires researchers to submit journal articles that arise from NIH funded research into the digital archive PubMed Central.¹³ This is a change from a voluntary to a mandatory policy and creates expectations of both investigators and institutions to ensure compliance. The offices that manage research funding will be on the front lines of ensuring that articles resulting from NIH funding are deposited. As grantees, institutions will be accountable for ensuring that authors retain the rights necessary to permit deposit of their works in PubMed Central (PMC). This will require authors to alter publication agreements in some cases.

NIH's process, though not complicated, requires navigation through an interface comprising 10 web pages and taking, on average (and according to NIH) 10 minutes per submission. Some institutions are developing a streamlined version, managed by the Library and employing a mix of technologies that include institutional repositories.

Development of such a system in tandem with an institutional repository would:

- * eliminate the need for our researchers to deal directly with the NIH web forms,
- * significantly reduce researcher time spent on compliance, and
- * ensure compliance for the institution as a whole.

In the days ahead, NIH will be providing additional information to assist institutions in meeting the new requirement.¹⁴

Tyler Walters states that

...with rapidly changing technologies, users now desire and expect transportable content that can be utilized within various digital environments and reused in multiple formats, and they need forums for the rapid exchange of ideas with both on-campus and external communities. In response, universities and the libraries hosting IRs are looking for ways to weave their repositories into the "information fabric" of their campuses' academic and business processes and catalyze changes in scholarly communications more broadly."¹⁵

IRs include benefits of expanded access, increased visibility and discovery, and support for open access to information. Researchers are finding that IRs can also provide a

¹³ See PubMed Central at <http://www.pubmedcentral.nih.gov/>.

¹⁴ NIH is maintaining an FAQ on the NIH Public Access Policy site at <http://publicaccess.nih.gov/>

¹⁵ Walters, Tyler O., "Strategies and Frameworks for Institutional Repositories and the New Support Infrastructure for Scholarly Communications," *D-Lib Magazine*, 12:10 (October 2006).
<http://www.dlib.org/dlib/october06/walters/10walters.html>

location for the sharing of data files that often are not included or supported by commercial publication.

IRs usually include a goal to support the long-term digital preservation of content. The preservation of scholarly communication requires rights management systems to address author rights and institutional copyright policies. Many authors are unclear about what intellectual property rights they have in general, and they may not negotiate their publishing contracts and therefore give away their copyrights. Researchers need to be encouraged to retain their right to deposit their work into an institutional repository.

During the past year, institutions and organizations have begun adopting author rights addendums. Author rights addendums address what rights an author retains. Retention clauses can be attached to publisher contracts. In turn, author rights and changes to scholarly communication need to be supported by the university. Currently, faculty use their own web pages, their department's web page, or a discipline-based depository to disseminate some version of their research when in fact they may not even own the copyrights to the published research. Peter Hirtle notes in his article "An Examination of Five Alternatives" that "unless addressed in the transfer agreement, the publisher may forbid an author to do the following:

- Post the work to the author's own web site, an institutional repository, or a subject-based repository.
- Copy the work for distribution to students.
- Use the work as the basis for future articles or other works.
- Give permission for the work to be used in a course at the author's institution.
- Grant permission to faculty and students at other universities to use the material.¹⁶

It is important for UK as an academic institution to provide the support required by researchers to ensure they have the rights to allow open access to research, knowledge, and creative ideas both internally but also to the wider community. Equally, it is important to build a sustainable repository that ensures long-term preservation of the University's digital resources.

Recommendation

Encourage University of Kentucky scholars to retain their rights to deposit their scholarship into an institutional repository.

Electronic Journal Publishing

As part of our Work Group charge, we were asked to define the issues surrounding e-journal content as they relate to an institutional repository. An institutional repository by

¹⁶ Hirtle, Peter B., "Author Addenda: An Examination of Five Alternatives," *D-Lib Magazine*, 12:11 (November 2006). <http://www.dlib.org/dlib/november06/hirtle/11hirtle.html>

definition supports content created by the institution. Most journal content is not institution-centric, so including e-journals seems beyond the scope of an IR. The impetus for this part of the Work Group charge comes from a specific need within the UK Libraries to develop a mechanism to preserve and distribute content we have purchased from publishers who are no longer able to manage content delivery on their own servers, i.e., when a publisher goes out of business. Journal backfiles from publishers in this situation are sometimes delivered to libraries to manage. Libraries must develop a way to preserve the content and allow access by patrons.

Digital journal content is not (generally) unique. If a publisher releases backfiles on CD or some other manner, that content is equally distributed to several universities and other subscribers. Intentional or accidental redundancy ensures the likelihood of persistence. Conversely, institutional repository content is at an acute preservation risk because it is generally not available in any other way. While e-journal backfile content is still at risk, most likely due to unsupported file formats and obsolete storage media, the duplicate availability mitigates the immediate acute risk.

We believe that e-journal backfile content has a place within the structure of the broader DOR, but that it should reside outside of the institutional repository. The content is not (necessarily) created by university-affiliated authors, nor is the content from a university-affiliated publication. However, the software available to serve university-produced content could easily be used to serve non-university-produced content. This content could fall within the scope of the DOR. The addition of this specific type of content to the DOR creates some complications that must be addressed:

- *Additional licensing/access restrictions.* License agreements may restrict access to university affiliates. The degree of complication depends on whether we have other access restrictions for the content with the DOR.
- *Possible format issues.* We may have access to the distributable format only, not an editable format, which may limit our ability to migrate or manage format-related access problems.
- *Data organization.* The content delivered by the publisher may contain little or no metadata, making it difficult to organize.

Other solutions are being developed in response to concerns about long-term access and preservation of e-journal backfile content. The most notable among those systems include:

- LOCKSS (<http://www.lockss.org/>)
LOCKSS is a system for distributed and redundant archiving of journal content. Multiple cache sites harvest and store journal content. Content in the distributed systems is compared for accuracy. Users access the cache only if the original publisher site is unavailable. Content is entered into the cache sites directly from the publisher site and cannot be loaded directly from CD or other mechanism.
- Portico (<http://www.portico.org/>)
Portico also relies on publisher agreements to harvest content, but storage is centralized on Portico servers rather than on distributed cache sites. Similar to LOCKSS, content cannot be loaded into Portico from backfiles delivered to libraries.

These solutions may be better options for dealing with e-journal backfiles in the future, but they do not currently address the immediate problem of dealing with content that has already been delivered to libraries for them to manage on a title by title basis.

Recommendation

Develop a University of Kentucky system to manage, preserve and deliver e-journal backfile content outside of the institutional repository but within the larger DOR.

Infrastructure and Resources Needed

There will be a shared responsibility for the actual development and continued maintenance of the DOR. The vision for the DOR features the University Libraries and UK Information Technology driving the development as well as the long term maintenance and sustainability of the system and preservation of the content. Full participation of a faculty advisory group will also be necessary to ensure the actualized system meets the needs of the academic community both internal and external to the University. Specific tasks will be designated to individual units while some areas of responsibility will overlap and will require collaborative management.

The Work Group envisions that the following University entities or communities will help build and sustain the DOR:

- University Libraries
- UK Information Technology
- Faculty Advisory Group
- Office of Research
- Office of the Provost
- University faculty, staff, and students

In concert with faculty advisors, the **University Libraries** will take the lead in:

- Program administration
- Development of policies and standards guiding acquisition and retention of content
- Development of the user interface
- Marketing and advocacy
- Development of user documentation
- Rights management
- Ingest of content
- Content recruitment
- User assistance in searching
- Participation in the creation of metadata

UK Information Technology will be responsible for:

- Infrastructure including hardware and software systems
- Data store and tools
- Documentation for all of the systems
- General system integrity including disaster recovery planning and security
- Plans for hardware refresh cycles

Together the **Libraries and UK Information Technology** will be responsible for developing and implementing the details of data standards as they impact migration and renewal of content.

In addition to marketing and advocacy, the **Faculty Advisory Group** will provide guidance and feedback on virtually all aspects of DOR program development and maintenance.

The **Office of Research** will help shape policies regarding deposit of research output into the DOR. The Office of Research will provide expertise regarding rights management related to funding agency requirements for access to grant-funded scholarship.

The establishment of the DOR must originate from the **Office of the Provost**. Because the DOR captures the scholarly output of the University, overall support must initiate from the University Administration. Support from the highest level will ensure compliance and participation by all colleges and research centers.

Fundamental to the success and the life of the DOR is the active participation of the University's faculty, staff, and students through the contribution of content. Therefore, the DOR features active and strong collaboration among all stakeholders, from the essential point of creation by University scholars to the support of DOR system management and preservation by UK Information Technology and UK Libraries. Finally, the success of the DOR depends on the full support of the University Administration.

Recommendation

It is anticipated that the resources needed to plan, develop, implement, and sustain the DOR will be substantial and will be more than any one campus unit can bear. Funding needs to be allocated centrally as the DOR will benefit the entire University enterprise. Needed resources include approximately six new staff positions allocated to the University Libraries and UK Information Technology. Additionally a substantial recurring hardware investment will be needed to support the endeavor. Given the importance of the content sustained and made accessible through this resource—the scholarly assets of the University—this represents a modest investment.

Summary of Recommendations with Suggested Action Steps

1. Without a central repository for digital objects, content creators are left to store and distribute their data using whatever means available to them at the time of creation. Given the fragility of digital content and the value of these digital assets to UK, the Work Group recommends that a digital object repository be established. The DOR will include an IR for faculty, staff, and student scholarship and will absorb and manage the digitized content currently produced and stored disparately throughout the University.
 - Identify a design and implementation team.
 - Establish standards.
 - Establish the responsibilities of the DOR Management Team/Service.
 - Design the system.
 - Develop a timeline for implementation.

2. The resources needed to plan, develop, implement and sustain the DOR will be substantial and will total more than any one campus unit can bear. Funding needs to be allocated centrally as the DOR will benefit the entire University enterprise. Needed resources include approximately six new staff positions allocated to the University Libraries and UK Information Technology. Additionally, a substantial recurring hardware investment will be needed to support the endeavor. Given the importance of the content sustained and made accessible through this resource—the scholarly assets of the University—this represents a modest investment.
 - Identify and allocate initial and recurring financial support.
 - Establish organizational structure for support of the DOR through the Office of the Provost.

3. Fundamental to the success and the life of the DOR system is the active participation of the University’s faculty through the contribution of content. Therefore active and strong collaboration among all stakeholders is essential.
 - Develop a marketing, outreach, and education program to sell the DOR.
 - Develop tools and documentation that enable ease of use.
 - Create a website to serve as the information and entry portal to the DOR.
 - Work with content creators to deposit digital content.
 - Work with departments to migrate existing collections of digital content.
 - Identify the “trustworthiness” of digital content and inform content creators of this status.

4. To adequately plan for a successful DOR, a survey of faculty research will inform the DOR’s development.
 - Conduct a University-wide survey to determine the types of digital objects being created and maintained.
 - Identify file formats.

- Research and analyze the “trustworthiness” of file formats.
5. Encourage University scholars to retain their right to deposit their scholarship into an institutional repository.
 - Develop a rights management education program.
 - Identify resources to support rights management.
 - Review and make known the institutional copyright policy.
 6. Develop a University of Kentucky system to manage e-journal backfile content outside of the institutional repository but managed and delivered through the larger DOR.
 - Establish protocol and standards for the ingest of digital content not created at the University.
 7. Stewardship of the digital assets of the University and the Commonwealth requires visionary and rigorous management and substantial resources. Without that that robust and proactive preservation administration, the University’s digital assets are rendered imperiled, if not totally unsustainable.
 - Establish a digital repository in compliance with the OAIS reference model.
 - Adopt the TRAC as the audit mechanism to ensure and document the repository’s trustworthiness.
 - Form an administrative body of digital stewards to conduct annual audits according to the TRAC checklist.

Conclusion

The Institutional Repository Work Group examined the issues related to digital content creation in a research environment and concluded that the University will benefit from ongoing, sustained access through a digital object repository. The DOR will function as a system to include current, disparate digital object repositories and will create the opportunity for deposition and maintenance of new collections. The institutional repository, a subset of the DOR, will offer faculty, staff, and students a vehicle, methodology, and a “safe place” for their scholarly efforts. The overarching DOR will incorporate the IR, e-journal content, digitized resources from the Libraries’ collections, and other digital resources created at the University. Therefore, shared support is appropriate, standards are required, and collaborative management is essential. The development, maintenance, and preservation of the DOR will require campus-wide administrative support with specific tasks and budgets assigned to both the University Libraries and UK Information Technology. This trusted digital repository will stimulate and encourage the creation of open scholarship and will ensure the persistence and usability of the University’s priceless digital assets.

Bibliography

- American Library Association. *Principles and Strategies for the Reform of Scholarly Communication*. 2006.
<http://www.ala.org/ala/acrl/acrlpubs/whitepapers/principlesstrategies.cfm>
(accessed November 16, 2007).
- Association for Library Collections & Technical Services (ALCTS) of the American Library Association (ALA) Preservation and Reformatting Section (PARS).
“Revisions for Discussion at ALA Annual 2007.” *Defining Digital Preservation*, July 6, 2007. <http://blogs.ala.org/digipres.php> (accessed November 16, 2007).
- “Association of Research Libraries :: Appendix II: Definition and Purposes of a Digital Library.” <http://www.arl.org/resources/pubs/mmproceedings/126mmappen2>
(accessed November 16, 2007).
- “Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources.” August 2001.
<http://www.oclc.org/programs/ourwork/past/trustedrep/attributes01.pdf> (accessed November 16, 2007).
- Center for Research Libraries. “CRL - TRAC Checklist & Report.” *Auditing and Certification of Digital Archives*, 2007.
<http://www.crl.edu/content.asp?11=13&12=58&13=162&14=91> (accessed November 27, 2007).
- “Digital Preservation Tutorial.”
<http://www.library.cornell.edu/iris/tutorial/dpm/foundation/oais/index.html>
(accessed November 16, 2007).
- Digital Repository Certification*. RLG and NARA, n.d.
<http://www.oclc.org/programs/ourwork/past/repositorycert.htm> (accessed November 16, 2007).
- Hirtle, Peter B. “Author Addenda: An Examination of Five Alternatives.” *D-Lib Magazine* 12, no. 11 (November 2006).
<http://www.dlib.org/dlib/november06/hirtle/11hirtle.html>. (accessed November 16, 2007).
- Lynch, Clifford A. “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age.” *ARL Bimonthly Report*, no. 226 (February 2003).
<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. (accessed November 16, 2007).
- “Reference Model for an Open Archival Information System (OAIS).” January 2002.
<http://public.ccsds.org/publications/archive/650x0b1.pdf> (accessed November 16,

2007).

“Trustworthy Repositories Audit & Certification: Criteria and Checklist.” February 2007.
<http://www.crl.edu/PDF/trac.pdf> (accessed November 16, 2007).

University of Kentucky Strategic Plan 2006-09. University of Kentucky, n.d.
<http://www.uky.edu/ucapp/plan.htm> (accessed November 16, 2007).

Walters, Tyler . “Strategies and Frameworks for Institutional Repositories and the New Support Infrastructure for Scholarly Communications.” *D-Lib Magazine* 12, no. 10 (October 2006). <http://www.dlib.org/dlib/october06/walters/10walters.html> (accessed December 8, 2006).

Appendix A: Work Group Charge

Institutional Repository Task Force – June 4, 2007

Charge – to provide a white paper which defines the concept of a University of Kentucky institutional repository and to outline the steps needed to move forward. The Task Force’s role is exploratory and is intended to create an understanding of the Library’s role in addressing the management of the University’s digital content. The task force will produce a report by October that will document:

1. the types of UK generated digital content to be potentially included in a repository and/or digital asset management system
2. a definition/understanding of the following in the UK context:
 - a. Institutional Repository
 - b. Digital Object Library
 - c. Scholarly communication issues that are related to repositories
 - d. Digital preservation
 - e. Relationship of the issues surrounding e-journal content
3. an analysis of the roles necessary to build and maintain an institutional repository including which of these the Libraries should potentially manage and which require responsibility on the part of other campus units.
4. specific recommendations for next steps

Membership – Initial membership of this task force is intentionally small to facilitate the writing of a white paper that addresses the “what” and “how” of a UK institutional repository. Key individuals and knowledge holders in various areas will be consulted and brought into the process as needed.

Members of the task force:

Lisa Carter, co-chair (note: Lisa left the University in August)

Mary Molinaro, co-chair

Brian Helm

Kathryn Lybarger

Beth Kraemer

Rebecca Ryder

Mary Beth Thomson

Deirdre Scaggs (joined Work Group in August)

John Soward, UK Information Technology

Appendix B: Audit Checklists

Example pages from various sections (3 checklists of 78)

Example from Section A: Organizational Infrastructure

Trustworthy Repositories Audit & Certification: Criteria Checklist					
Organization:		Auditor:		Page	
Section:	A. Organizational Infrastructure	Interviewee(s):		Date	
Aspect:	A3. Procedural accountability & policy framework				
Criterion	Evidence (Documents) Examined	Findings and Observations		Result	
A3.1. Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.					
A3.2. Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.					
A3.3. Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.					
A3.4. Repository is committed to formal,					

periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.			
A3.5. Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.			

Example from Section B: Digital Object Management

Trustworthy Repositories Audit & Certification: Criteria Checklist				
Organization:		Auditor:		Page
Section:	B. Digital Object Management	Interviewee(s):		Date
Aspect:	B.2 Ingest: creation of the archivable package			
Criterion	Evidence (Documents) Examined	Findings and Observations	Result	
B2.6. If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).				
B2.7. Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).				

B2.8 Repository records/registers Representation Information (including formats) ingested.			
B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information			
B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.			

Example from Section C: Technologies, Technical Infrastructure & Security

Trustworthy Repositories Audit & Certification: Criteria Checklist					
Organization:		Auditor:		Page	
Section:	C. Technologies, Technical Infrastructure & Security	Interviewee(s):		Date	
Aspect:	B.2 System Infrastructure				
Criterion	Evidence (Documents) Examined	Findings and Observations		Result	
C1.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.					
C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).					
C1.8 Repository has a documented change management process that identifies changes to critical processes that					

potentially affect the repository's ability to comply with its mandatory responsibilities..			
C1.9 Repository has a process for testing the effect of critical changes to the system.			
C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.			

Appendix C: Sample Faculty Survey

Methodology: In the implementation phase of the DOR, a Work Group comprising some members of UK Libraries and UK Information Technology will conduct an online survey of University researchers to determine what digital objects they currently create and attempt to manage. Promotion and advertisement of the survey and its intent will precede the survey. The results will be tabulated, analyzed and evaluated by the Work Group. A report of findings will be issued. The results will form the basis of DOR and trusted digital repository development.

Sample Questions:

1. What types of electronic teaching materials do you create? Where applicable, specify format (e.g., PDF, Microsoft Word, LaTeX)
 - a. class slides (PowerPoint/Beamer)
 - b. textbooks
 - c. course notes
 - d. assignments/quizzes/tests
 - e. learning modules (SCORM, WebClass)
 - f. web sites
 - g. multimedia
 - h. other (specify)

2. How do you share these with your colleagues?
 - a. email
 - b. web site
 - c. other

3. What types of digital materials do you create for publication?
 - a. papers, technical reports
 - b. books
 - c. images
 - d. audio files
 - e. video files
 - f. multimedia
 - g. computer programs
 - h. data sets
 - i. other (specify)

4. How do you communicate/share your finished work to your colleagues?
 - a. alert them to the journal it will be in
 - b. email
 - c. web site
 - d. postal mail
 - e. public preprint archive (such as arXiv)
 - f. other

5. Do you typically retain the rights to material that you publish?
6. Have you ever used an author rights addendum?
7. Do you create materials too large/complex to include with your published work?
If so, what type?
 - a. data sets
 - b. computer programs
 - c. simulations
8. How do you stay abreast of the research of your colleagues at UK?
 - a. word of mouth?
 - b. department meetings?
 - c. department newsletter/listserv?
 - d. periodical review (such as PubMed, MathSciNet)
9. Do you have a website?
 - a. on a University/department server?
 - b. with a home internet provider
 - c. Do you update it regularly?
10. Have you ever lost data due to
 - a. hardware failure
 - b. file corruption
 - c. human error (accidentally erasing it)
 - d. lack of organization (losing track of where it is)
 - e. hardware/software/format obsolescence
11. Have you ever used an institutional repository at another institution? If so, do you have any comments on it?
12. Have you ever retrieved a copy of a colleague's work from their institutional repository?

Appendix D: Other Resources Consulted

- About Deep Blue*. “Deep Blue is the University of Michigan’s permanent, safe, and accessible service for representing our rich intellectual community.”
<http://deepblue.lib.umich.edu/about/index.html>
- Association of Research Libraries. *SPEC Kit 292 Institutional Repositories*. 2006.
- Davis, Philip M. and Matthew Connolly. “Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace,” *D-Lib Magazine* Vol. 13 no. 3/4 (March/April 2007) p. 1. <http://www.dlib.org/dlib/march07/davis/03davis.html>
- Gibbons, Susan. Establishing an Institutional Repository [Special issue]. *Library Technology Reports* Vol. 40 no. 4 (July/August 2004) p. 5-67.
- Green, Ann and Myron P. Gutmann, “Building partnerships among social science researchers, institution-based repositories and domain specific data archives,” *OCLC Systems & Services* Vol. 23 no. 1 (2007) p. 35-53
- Hixson, Carol. “If We Build It, Will They Come (Eventually)? Scholarly Communication and Institutional Repositories,” *The Serials Librarian* Vol. 50 no. 1/2 (2006) p. 197-209.
- Hoorens, Stijn, Jeff Rothenberg, Constantijn van Oranje, Martijn van der Mandele and Ruth Levitt. *Addressing the Uncertain Future of Preserving the Past*. Santa Monica, CA: The Rand Corporation, 2007. http://www.rand.org/pubs/technical_reports/TR510/
- John, Nancy. “Digital Repositories: Not Quite at Your Fingertips,” *Libri*, Vol. 55, (2005) pp 181-107. <http://www.librijournal.org/pdf/2005-4pp181-197.pdf>
- Johnston, Leslie. “An overview of digital library repository development at the University of Virginia Library,” *OCLC Systems & Services* Vol. 20 no. 4 (2004) p. 170-3.
- Lowry, Charles B. “ETDs and Digital Repositories--a Disciplinary Challenge to Open Access?”. *portal: Libraries and the Academy*, Vol. 6 no. 4 (October 2006), p. 387-93.
http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v006/6.4lowry02.pdf
- Markey, Karen et al. *Census of Institutional Repositories in the United States: MIRACLE Project Research Findings*. Washington, DC: Council on Library and Information Resources. (February 2007) <http://www.clir.org/pubs/reports/pub140/pub140.pdf>
- Nolan, Christopher W. and Jane Costanza. “Promoting and Archiving Student Work through an Institutional Repository: Trinity University, LASR, and the Digital Commons,” *Serials Review* Vol. 32 no. 2 (June 2006) p. 92-8.
- PADI: Preserving Access to Digital Information*. “PADI is a subject gateway to international digital preservation resources National Library of Australia.”
<http://www.nla.gov.au/padi/>.

Rogers, Sally A. "Developing an Institutional Knowledge Bank at Ohio State University: From Concept to Action Plan," *Portal: Libraries and the Academy*, Vol. 3, No. 1 (2003), pp. 125–136.

http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v003/3.1rogers.pdf

UR Research at the University of Rochester. "The University of Rochester Libraries use UR Research as a long-term storage system for digital works; it is a digital repository designed to capture, store, index, distribute, and preserve the intellectual output of a university's faculty in digital format.

<http://www.library.rochester.edu/index.cfm?page=1346&CFID=11396929&CFTOKEN=57155047&jsessionid=5a302514cdee193022d2>

Appendix F: Glossary

Administrative metadata - information necessary to allow a repository to manage the object: this can include information on how it was scanned, its storage format etc (often called *technical metadata*), copyright and licensing information, and information necessary for the long-term preservation of the digital objects (*preservation metadata*)

AIP (Archival Information Package) - An AIP is the digital equivalent of an archival item such as a book, a record album, or a motion picture. It consists of multiple data files that contain the digitized content of the archival item. In addition to the data files, the AIP contains metadata that describes the structure, content, and meaning of the data files. The data files and metadata are packaged (encapsulated) either logically or physically as an entity. AIPs are used to transmit and/or store archival objects within a digital repository system.

ANSI (American National Standards Institute) - The Institute oversees the creation, promulgation and use of thousands of norms and guidelines that directly impact businesses in nearly every sector: from acoustical devices to construction equipment, from dairy and livestock production to energy distribution, and many more. ANSI is also actively engaged in accrediting programs that assess conformance to standards—including globally-recognized cross-sector programs such as the ISO 9000 (quality) and ISO 14000 (environmental) management systems.

arXiv - “a highly-automated electronic archive and distribution server for research articles” (<http://arxiv.org/help/general>) arXiv is a common repository for disseminating preprints in scientific fields of study.

Author rights addendum - a document which an author may submit to a publisher as an addendum to their standard publication agreement which, if accepted by the publisher, allows the author to retain some rights regarding the re-use and dissemination of their work. One example is the SPARC Author Addendum (<http://www.arl.org/sparc/author/addendum.html>)

Center for Research Libraries (CRL) - “CRL is a consortium of North American universities, colleges, and independent research libraries. The consortium acquires and preserves traditional and digital resources for research and teaching and makes them available to member institutions through interlibrary loan and electronic delivery.” (<http://www.crl.edu/>)

Checksum - a small file created by performing some algorithm on a larger file to allow for future detection of errors. Popular checksums include file size, CRC-32 ("cksum") and MD5.

Consultative Committee for Space Data Systems (CCSDS) - “a multi-national forum for the discussion of common space communications issues”.
(<http://public.ccsds.org/default.aspx>)

Content Management System (CMS) - a system used for organizing, managing and disseminating digital content. Popular content management systems include Fedora, DSpace, DLXS, DAITS, and CONTENTdm.

Dark Archive - an archive that is not accessible to the public. A dark archive may be restricted to use by certain individuals or organizations, but the term may also refer to an archive that cannot be crawled by commercial search engines such as Google.

Derivative - a file created from another file, often for some specific purpose. Smaller derivative files may be created from the larger preservation master files (such as PDF or JPG from TIFF, or MP3 from WAV) for ease of distribution.

Descriptive Metadata - metadata that describes the data and allows for searching or identifying of the data.

Digital Object - an item as stored in a digital library, consisting of data, metadata, and an identifier.

DIP (Dissemination Information Package) - a derivative of AIPs delivered to the consumer of a TDR. These files are often different from (and smaller than) SIPs and AIPs.

DLXS (Digital Library eXtension Service) - a content management system developed and maintained by the University of Michigan. Kentuckiana Digital Library uses DLXS as a platform.

Granularity - the level at which metadata is assigned. Metadata assigned to a newspaper title, an issue, and a page would be at different granularities.

Interoperability - a computer hardware or software system's capability to exchange data effectively with a different type of system and use the shared data.

Institutional repository (IR) - an online locus for collecting, preserving, and disseminating—in digital form—the intellectual output of an institution, particularly a research institution.

KDL (Kentuckiana Digital Library) - repository for UK and other statewide institutions' multi-image collections hosted by the Kentucky Virtual Library

LE (Life Expectancy) - the probable number of years that information as recorded on various media is usable. Factors affecting LE include temperature, humidity, media composition and chemistry, physical condition, software and hardware obsolescence and data corruption.

LaTeX - a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation. LaTeX is the *de facto* standard for the communication and publication of scientific documents.

Learning Module - a well-defined learning unit, which consists of a hierarchic structure and associated materials.

Metadata - data about data. See also **administrative metadata, descriptive metadata, preservation metadata, rights and access metadata, structural metadata**

OCLC - “OCLC Online Computer Library Center is a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing information costs.”

(<http://www.oclc.org/about/default.htm>)

Open Access (OA) - a publication model where in neither readers nor a reader's institution are charged for access to articles or other resources. Users are free to read, download, copy, distribute, print, search, or link to the full texts of these articles.

OAI-PMH (Open Archives Initiative's Protocol for Metadata Harvesting) - a protocol created to facilitate discovery of resources distributed in many repositories or locations. The OAI-PMH achieves this by providing a simple, yet powerful framework for metadata harvesting. Harvesters can incrementally gather records contained in OAI-PMH repositories and use them to create services covering the content of several repositories.

Open Archival Information System (OAIS) - an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS archive to be distinguished from other uses of the term ‘archive’. The term ‘Open’ in OAIS is used to imply that standards are developed in open forums, and it does not imply that access to the archive is unrestricted

Open Source - refers to any program whose source code is made available for use or modification as users or other developers see fit.

PDF (Portable Document Format) - the native file format for Adobe Systems' Acrobat. PDF is the file format for representing documents in a manner that is independent of the original application software, hardware, and operating system used to create those documents. A PDF file can describe documents containing any combination of text, graphics, and images in a device-independent and resolution independent format. These documents can be one page or thousands of pages, very simple or extremely complex with a rich use of fonts, graphics, color, and images.

PREMIS - XML standards from the PREMIS (PREservation Metadata: Implementation Strategies) Working Group to record information about preserving data.

Preservation Metadata - metadata that aids in the preservation of data. May include checksums and information about the format of the data.

Proprietary - something that is used, produced, or marketed under exclusive legal right of the inventor or maker

RLG (Research Libraries Group) - "A not-for-profit membership corporation of institutions devoted to improving access to information that supports research and learning." RLG has merged with OCLC.

Rights and Access Metadata - metadata that describes how (and by whom) data may be accessed, used and distributed.

SIP (Submission Information Package) - A Submission Information Package, commonly referred to as a "SIP," is created by a Producer to prepare records for transfer to an Archive. OAIS defines a SIP as "an Information Package that is delivered by the Producer to the OAIS [Preservation System] for use in the construction of one or more AIPs."

Structural Metadata - metadata that describes how data objects are related, or how they can be combined to form a compound object. For example, structural metadata might describe how newspaper pages combine to form sections or issues of a newspaper.

Trusted Digital Repository - a repository with the mission to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.