Institutional Repositories:
Enhancing Teaching, Learning, and Research

EDUCAUSE Evolving Technologies Committee
Alan McCord, Lawrence Technological University / University of Michigan
October 16, 2003

Introduction

Our institutions are deploying increasingly complex educational storage and delivery systems ranging from course management systems, personalized portals, student portfolio systems, streaming media services, and Web content management systems. Meanwhile, we have already made substantial investments in integrated library systems and enterprise resource planning systems, and in the underlying IT infrastructure to manage these services.

These major institutional systems are integrated to various degrees, but do not yet provide the type of seamless integration that can leverage new approaches to teaching, research, and administration. Furthermore, our ability to capture and creatively use rich media objects in our present environment lags behind our ability to capture and use data and still images. Finally, we lack a structured approach – both technological and procedural – to capture and preserve the intellectual output of the campus community for future access.

What Is An Institutional Repository?

An institutional repository consists of formally organized and managed collections of digital content generated by faculty, staff, and students at an institution. The content of these repositories can be available for integration with on-campus library and course management systems, and can also be made available to colleagues and students at other institutions, as well as to the general public.

When we use the term “repositories” today, we can be speaking about one of many different technologies that support the storage and distribution of digital content, including:

1. Collection-based digital repositories managed by library professionals, either stand-alone or aggregated;
2. Course management systems and associated file stores;
3. Collections of research data and reports managed by academic departments;
4. Student academic portfolio systems;
5. Institutional file storage systems;
6. Digital asset management workflow systems; or

7. Web content management systems used by institutions or departments to store and stage Web content.

While many of these components can play roles in capturing and managing digital content, an institutional repository is a more specific concept – a centrally managed collection of institutionally-generated digital objects designed to be maintained in perpetuity. An institutional repository will be capable of indexing and serving a wide range of static and moving images, and will be seamlessly visible from course management systems, integrated library systems, administrative workflow systems, and via public portals. Institutional repositories can be viewed “as a natural extension of academic institutions’ responsibility as generators of primary research seeking to preserve and leverage their constituents’ intellectual assets; and as one potentially major component in the evolving structure of scholarly communication.”

---

**Why Are Institutional Repositories Important to Higher Education?**

Students and faculty members increasingly recognize the need to store their intellectual output in the form of personal collections, and to make available the results of their work within and outside the institution. Institutions increasingly recognize the need to develop repositories of intellectual output for long-term archival purposes, and to administer the property rights associated with stored assets.

We presently face a great disparity between our ability to create digital assets and our ability to control our inventories of digital assets. Our libraries have done an outstanding job of building controlled asset inventories of print materials, and increasingly of electronic journal content form within and outside the institution. Most of the collections managed by our libraries, however, are created outside the walls of our institutions. Many newer electronic journal collections are inadequately integrated with the master library catalog.

Our institutions have done a less adequate job, however, of providing the tools to create, manage, and inventory our rich media assets (videos, audios, presentation materials, datasets, records of collaborations, research works in progress) as well as more traditional digital assets created by our faculty and students. Our libraries have largely determined the contents of the collections they choose to manage on behalf of the institution. We have somewhat haphazardly built an inventory of our internally generated intellectual property without providing the inventory control needed to provide access.

It is arguably this internally generated intellectual property – much of which does not migrate to scholarly journals – that should be of greatest importance to the future of our institutions. Would we tolerate the elimination of our libraries’ existing asset management systems? Of course we should not – so why would we perpetuate a situation in which most of the assets we create are not managed? What value are we losing by not adequately managing our intellectual output? What value can we gain by being able to catalog, identify, and re-purpose the vast amount of knowledge that has been generated by our faculty and students?

The potential of institutional repositories to help foster change within the academy is significant. Clifford Lynch points out that perhaps the most important potential payoff of institutional repositories is “opening up entire new forms of scholarly communication that will need to be legitimized and nurtured with guarantees of both short- and long-term accessibility.”

---

What Are The Implementation Challenges?

As with most change programs, the most significant challenge facing institutional repositories is the "administrative attention span" and long-term commitment to insure preservation and maintenance of the repository over time, providing the necessary confidence to enable faculty members to contribute their works to the repository.

Issues associated with the asset management focus on developing metadata standards for digital objects, establishing long-term preservation approaches for digital assets, and developing collection management processes. Assets contributed to a personal or departmental repository may migrate over time into the institutional repository, so their "life cycles" must be considered as asset management practices are designed. Some digital assets will possess "multiple concurrent life cycles" – for example, a digital video clip being used by a department for course delivery while the entire digital asset is contributed to and managed through the institutional repository. "Decision-making both on digital conversion of analog material and on metadata assignment needs to be carefully planned at the start of a digital project."

IT infrastructure issues include the impact of successful repositories on institutional storage and network architectures. In this regard, our network infrastructures are generally able to carry the type of traffic served by an institutional repository, especially with deployment of multi-cast capabilities and end-to-end performance management. Our storage architectures require more attention: we frankly need to increase our shared capacities by orders of magnitude while at the same time reducing our storage cost/unit/time by orders of magnitude – a "two tailed order of magnitude problem." And our authentication and access systems need to be improved to support role-based access to various renderings of digital assets.

We also need to improve our capabilities in the areas of rich media capture, ingestion and logging, providing ubiquitous and easy-to-use services to capture intellectual output that today is lost. At the same time, we need to identify how best to digitize the most important non-digital content for inclusion in the repository. We must also pay attention to diversity – of media types, of workstation platforms, of rendering methods – and the long-term preservation and/or migration issues associated with a commitment to diversity.

The most significant policy issue surrounding institutional repositories is the management of digital rights. The importance of digital rights for a particular repository object may change radically over time, and objects not exclusively in the public domain require the use of sophisticated rights management systems to protect and manage the distribution of those objects. Efforts by the Creative Commons to develop standards-based use licenses are addressing one important element of the digital rights issue. Current digital rights management systems focus largely on the fee-based distribution of copyrighted material using proprietary mechanisms, and are therefore not well suited to higher education.

Who Are The Major Institutional Repository Players?

There are a number of institutional repository initiatives underway within higher education. Some of the most visible initiatives include:

1. The Open Archives Initiative develops and promotes interoperability standards to facilitate the dissemination of content. Their Metadata Harvesting Protocol provides mechanisms for multiple disciplines to contribute to an institutional repository using common metadata. Participants include the

---

4 URL: http://www.creativecommons.org
2. MIT’s DSpace is an open source software platform that enables capture and submission of works, distribution of those works, and long-term preservation of assets. DSpace’ endeavors to create a federated collection of intellectual resources from the world’s leading research institutions.

3. Harvard’s Digital Repository Service (DRS) provides Harvard owners of digital material with a storage and retrieval system for their collections. Services and facilities include an electronic storage facility, management of administrative and structural metadata, preservation policies and procedures, and delivery of objects to front-end systems such as online catalogs.

4. The Flexible Extensible Digital Object and Repository Architecture (Fedora) is a foundation for developing interoperable digital libraries and institutional repositories using XML and Web services. Jointly developed by the University of Virginia and Cornell University, and sponsored by the Mellon Foundation.

5. The Canada ARL (CARL) Institutional Repositories Pilot Project is implementing institutional repositories at nine Canadian universities.

6. The University of Southampton’s EPrints initiative is designed to manage disciplinary or institutional print collections, rather than digital collections. Eprints software is OAI compliant and freely available under a GNU license, and is in use at California Institute of Technology, the University of Queensland, and other institutions.

Commercial activities generally focus not on institutional repositories proper, but rather on related technologies that leverage existing products in several spaces, including:

1. Library catalog systems, especially those using SFX context-sensitive reference linking to launch digital objects that are indexed within the catalog.

2. Academic journal and electronic content providers;

3. Digital asset management systems, especially those incorporating standards-based and flexible metadata and open APIs;

4. Portal solutions, especially those incorporating content management functionality; and

5. Student portfolio management systems, designed to provide a repository for student work for the purposes of supporting outcomes assessment or for recruiting.

Student portfolio work is also occurring in the higher education community. The University of Minnesota Open Source Portfolio Initiative (OSPI) is developing an open source electronic portfolio designed to maintain student output for life while interoperating with other knowledge management systems.
What Are The Rules of Thumb for Judging Among The Options?

Due to the wide variance of approaches in the development of institutional repositories, and the young state of the market, the rules of thumb relate to the nature, scope, and depth of your institutional vision for repositories:

1. The extent to which rich media will be integrated into course management and delivery systems;
2. The degree of integration between the institutional repository, library catalog systems, institutional portals, and other asset management systems;
3. The extent to which the institution can commit to long-term centralized storage of repository assets; and
4. The extent to which the institution can agree to develop metadata standards, asset life cycle definitions, and repository practices.

How Are Institutional Repositories Evolving?

Given the many aspects of the repository market space, and the many projects underway already, here is our assessment as to how the overall market space is evolving:

1. Integrated library systems are evolving to include rich media management and display capabilities, and the use of new metadata standards for digital objects;
2. Portal environments are maturing and are incorporating Web content management functionality;
3. Course management systems still have immature links to integrated library systems and to object repositories located outside its base architecture;
4. Most digital asset management implementations are still focused on delivering services to “operational silos” with standardized business practices (e.g. marketing, brand management, regulatory processes);
5. There are a number of institutional repository demonstration projects at the discipline and institutional level as described above. Capabilities for integrating multiple repositories, including metadata harvesting and federated searching across repositories, are being developed; and
6. There are as of yet no viable commercial institutional repository offerings due to the early state of the market, the complexity of repository requirements, and the small potential customer base.

When Will Institutional Repositories Become Essential on Campuses?

We expect that institutional repositories will soon become a significant focus of most higher education institutions within three years. Some of the key drivers that will determine the rate and depth of change include:

1. The deployment of low-cost and easy-to-use transcoding tools to re-purpose rich media to multiple formats, and for display in multiple environments (e.g. from wireless phones to tablet computers to classroom projection systems);
2. The success of demonstration and early production projects – using DSpace and similar tools – as repositories for storing faculty members’ scholarly work;
3. The availability of seamless integration between integrated library systems and course management systems, largely driven by the adoption of metadata standards and sharing of instructional objects;

4. The widespread use of consumer digital video recording and editing tools. This will result in commonplace digital video capture in classrooms by faculty members and students, and distribution of these assets over the network using peer-to-peer technologies;

5. The demand for “portable student portfolios” that will accompany students throughout their academic career, and the potential use of portfolios for admissions decision-making and job recruitment; and

6. The successful use of Web Content Management Systems to manage Web content for academic departments and libraries.\(^{13}\)

How Should We Proceed?

The first step of planning an institutional repository initiative is to “prepare key decision-makers to understand the importance of digital archiving of the scale needed to support scholarship.”\(^{14}\) Beyond that crucial step, here are suggested steps for developing an institutional repository:

1. Early and significant collaboration with library experts to design basic metadata standards and cataloging procedures;

2. Early decisions – which should be revisited later – regarding the vision and scope for an institutional repository, and for the services to be provided to faculty and students by the stewards/managers of the repository;

3. An early decision on whether to immediately implement a true repository, or rather to evolve related systems with the intent of integrating those systems into a repository environment at a later date;

4. Early “small wins” with individual academic departments focused on integrating institutional repository objects into classroom instruction or research projects;

5. Addressing underlying infrastructure issues, including network bandwidth, low-cost disk storage, authentication, and digital preservation; and

6. Develop institutional standards and procedures for managing disciplinary and institutional collections, and procedures for migrating content from individual to disciplinary to institutional status.

Conclusion

We are likely to see the concept of institutional repositories develop in both convergent and divergent ways over the next few years. Not every institution will develop a formally managed institutional repository along the lines of DSpace. But every institution that is utilizing course management systems, library catalog systems, and student portfolio systems will see increased “repository-like” functionality in their products. The open source movement, coupled with greater network collaboration among researchers, should give rise to discipline-specific federated repositories hosted by institutions, research projects, or professional associations.


Institutional repositories are a visible manifestation of the emerging importance of knowledge management within higher education. Paradoxically, “scholarly respect for knowledge and a desire to ensure academic freedom make most institutions reluctant to manage knowledge of any sort.” The long-term impact of institutional repositories is likely to change many of the basic assumptions about how intellectual output is managed by individuals, their colleagues, and the academy, and how research itself is conducted.

Related EDUCAUSE2003 Sessions

The following EDUCAUSE 2003 sessions relate directly or indirectly to the topic of institutional repositories.

1. MIT OpenCourseWare: A New Model for Open Sharing – Wednesday, November 5, 10:30 a.m. - 11:20 a.m., Ballroom A
2. Strategies for Sharing Learning Content and University Knowledge Across Departments and Campus Boundaries, Wednesday, November 5, 10:30 a.m. - 11:20 a.m., Room 210A
3. ProfPort Webfolio System: Implementation, Curriculum, and Assessment, Wednesday, November 5, 11:40 a.m. - 12:30 p.m., Room 210C
4. Introducing the Blackboard Content System, Wednesday, November 5, 2:15 p.m. - 3:05 p.m., Room 213D
5. Buy, Build, or Adapt an E-Portfolio System? The Benefits of Multiuniversity Collaboration, Wednesday, November 5, 3:50 p.m. - 4:40 p.m., Room 210A
6. Portfolios (Current Issues Roundtable), Wednesday, November 5, 4:55 p.m. - 6:10 p.m., Ballroom C
7. E-Learning Objects: The Value of SCORM and MPEG-7 Packaging for Digital Media Assets, Thursday, November 6, 11:45 a.m. - 12:35 p.m., Room 303B
8. Emerging Best Practices for Integrating Library Content and Services with Educational Technology, Thursday, November 6, 2:20 p.m. - 3:10 p.m., Room 210A
9. Development and Implementation of a SCORM-Compliant Course: Successes, Issues, and Lessons Learned, Thursday, November 6, 2:20 p.m. - 3:10 p.m., Room 303B
10. Designing an Electronic Portfolio System for a Large Research University in Asia, Thursday, November 6, 3:55 p.m. - 4:45 p.m., Room 204A
11. MIT OpenCourseWare: Beyond the Pilot (Poster Session), Thursday, November 6, 4:55 p.m. - 6:10 p.m., Exhibit Hall B, Table 29
12. Authority of Consensus: Next-Generation Course Management System Features, Friday, November 7, 8:10 a.m. - 9:00 a.m., Room 210A
13. Electronic Portfolios: The Triple Helix of Learning, Assessment, and Pedagogy, Friday, November 7, 8:10 a.m. - 9:00 a.m., Room 303D
14. A First Look at Electronic Portfolio Implementations, Friday, November 7, 9:30 a.m. - 10:20 a.m., Room 204A