DIGITAL LIBRARY
FEDERATION

Digital Library Content and Course Management Systems: Issues of Interoperation

Report of a study group funded by the Andrew W. Mellon Foundation

Co-Chairs: Dale Flecker, Associate Director for Planning & Systems, Harvard University Library, and Neil McLean, Director, IMS Australia

July, 2004

Table of Contents

Executive Summary

With funding from the Andrew W. Mellon Foundation, an *ad hoc* group of digital librarians, course management system developers, and publishers met under the aegis of the Digital Library Federation to discuss the issues related to the use of digital library content in course management systems. The size, heterogeneity, and complexity of the current information landscape create enormous challenges for the interoperation of information repositories and systems that support course instruction. The group has created a checklist of things that operators of digital content repositories can do to help ameliorate the complexities of such interoperation. It also explored through the means of use cases the utility of tools which help instructors gather information resources from various distributed information repositories for teaching purposes, and created a model of how the group envisions the interaction of users, tools, and information repositories in the future. Understanding the complexities of the information landscape, and the importance of tools to simplify interactions with that landscape, is critical for those building systems and services in this domain. The group believes that it is now important that the community move from theoretical discussions of interoperation of content repositories and instructional systems to real-world demonstration projects in order to further our collective understanding of the needs of users and the realities of systems interoperation.

Introduction

American institutions of higher education today are awash with digital information resources. Members of the educational community commonly have access to thousands and thousands of electronic books and journals, hundreds of digital reference works, increasingly rich collections of digital pictures, videos, and music, and large databases of survey, geographic, and scientific data. Few areas of academic work are not dependent on at least some digital resources at this point, and the range and importance of what is available continues to grow dramatically

While many digital resources are maintained and accessed through the local environments of scholars and research groups, a very significant number, particularly materials of wider interest, are captured in the more formal systems of publishers, digital libraries, and institutional repositories.

In the same period that the range and scale of digital resources available within universities was beginning to grow dramatically, so was the use of information technology tools to assist in or augment teaching and learning. These tools range from the small and personal (personal web sites, PowerPoint) to large-scale institutional course management systems. Given the richness of digital resources available, one might have expected that educational tools would quickly become a significant vehicle for providing students with access to digital library resources relevant to their courses. However, there is a wide-spread perception that the level of integration of digital materials from formal repositories in educational systems remains relatively low.

An awareness of the need for interoperation of repositories of quality content with systems supporting learning and teaching has been growing over the past few years. The issue has been addressed on a technical architecture level in the "repository interoperation" work of IMS and OKI. On a more immediate and short-term level, the IMS Digital Library Special Interest Group has a subgroup working on standards for the exchange of "resource lists" (structured lists of readings and similar materials) between course management systems, integrated library systems, and other related entities (resource lists containing pointers to digital resources represent one form of integration of digital resources into learning systems).

In order to further progress in this area, the Andrew W. Mellon provided support for an *ad hoc* group of digital librarians, course management system developers, and publishers to meet and discuss some useful next steps to increase the integration of existing digital resources into the working environments of instructors in higher education. The Group (see *Appendix A* for a list of participants), co-chaired by Dale Flecker of Harvard University and Neil McLean of IMS Australia, met face-to-face twice, in August and December, 2003. It spawned two working groups, each of which wrote a report, as discussed later. This paper summarizes the work of the Group as of March, 2004

General findings and observations

.

Given the breadth of the topic, it is unsurprising that discussions of the Group ranged over a large number of issues. Among the issues and observations that most affected the direction of the Group are:

- The barriers to finding and re-using extant digital materials in a course context are very high today. Locating an appropriate place to look for materials, finding individual resources within the systems in which they are described, reusing existing descriptive metadata, coping with access management systems, understanding technical formats and intellectual property constraints, and ensuring continuing access to selected objects are all difficult. We do not have systems in place which make it easy for instructors with limited time and very limited technical expertise to simply locate and reuse digital content.

- The universe of systems containing materials useful in teaching and learning is highly diverse. This diversity is a reflection of many factors: differing types of digital objects (geographic databases versus art images), different organizations (Harvard's collections versus MIT's), different technical sophistication (an e-journal published by Elsevier versus one published in an academic department), and different intent (legal information systems versus genomic databases). The combination of these factors represents a major challenge in creating a coherent and easily useable information environment for instructors assembling resources for a course.

- The universe of systems containing materials useful in teaching and learning is very large. It includes not just systems internal to universities (institutional repositories, digital libraries, learning object repositories, museums) but also a very large number of commercial and non-commercial publishers which are normally accessed through the library under licensing agreements.

- The diversity of players in the digital domain is an impediment to the wide-spread implementation of any proposed solutions to simplifying the environment. While there are important content holders that are part of the educational environment and who see the support of education as one of their important roles, there are many more players for whom educational use is a minor or incidental part of their business. As we identify measures that content players can take to simplify access and reuse, it will be important for the educational community to work with content holders to explain the importance of, and where possible, induce them to adopt, such measures.

- Simplifying the use of resources goes well beyond questions of repository access protocols and standards. The need to work with system operators and data owners was noted above. The scale of available resources will require considerable management effort in terms of selection, storage and organization. Intellectual property concerns will require efforts in licensing, vendor relations, and education. Heterogeneous systems interfaces and metadata practices will require efforts in data conversion, in the building of agents to cope with diverse interfaces, and in constant monitoring to cope with incessant change. These efforts will necessarily involve many parts of the university.

- Much of the thinking to date in this domain has concentrated on formal course management systems. But we know that instructors use many different kinds of systems to deliver teaching materials, including, for example, the use of PowerPoint for classroom presentation, or of a course weblog to share work.

- Tools and systems relevant to discovering and using information resources are being acquired or created elsewhere in the university, and their integration into the learning systems environment will take effort. Metasearch engines, capable of simultaneously searching multiple systems and of homogenizing retrieved metadata, are potentially key components of integration. OpenURL linking servers, capable of resolving metadata links to on-line or physical resources, are also logical parts of the integrating environment. It is worth noting the growing number of both free and commercial sources (e. g., Serial Solutions, SFX Knowledgebase, TD-Net, Jake) which maintain systematic information about electronic resources including up-to-date interface requirements. These are becoming important parts of digital library infrastructures, and could play an important role in integrating information

resources into course environments, as they help deal with the diversity and continual change in the information landscape.

During the first deliberations of the Group, David Greenbaum of the Interactive University project at Berkeley introduced a diagram (see *Figure 1*) which captures many of the points above and helped focus the thinking of the Group.

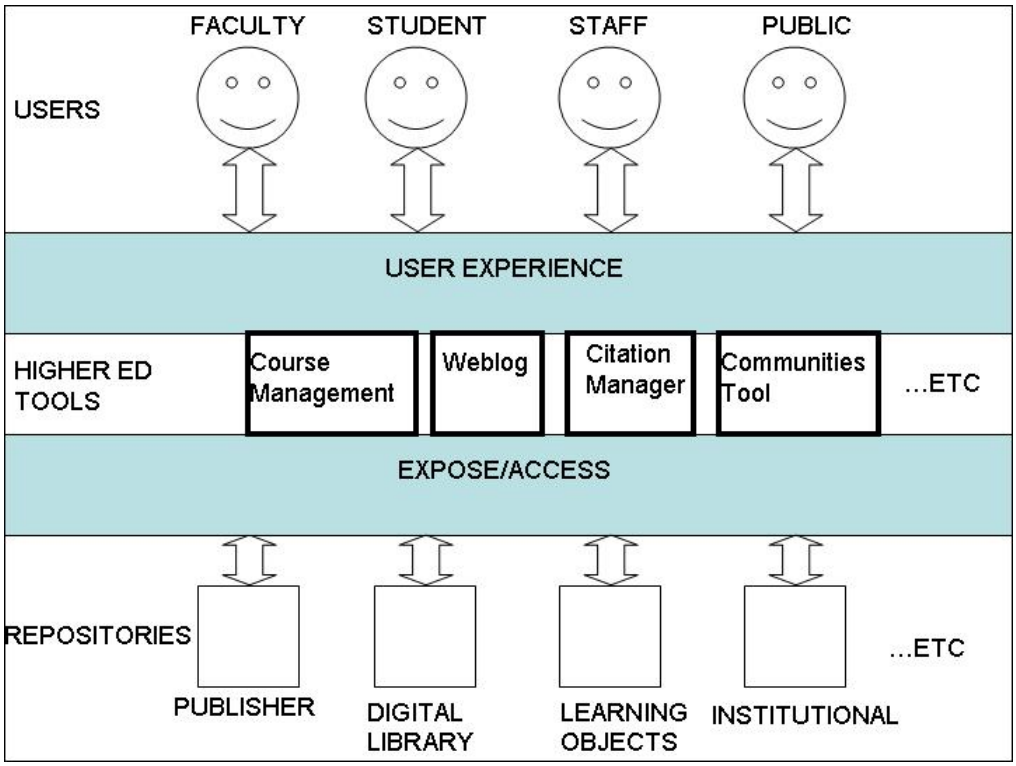

Figure 1. **Users, Tools, and Repositories**

The diagram posits that diversity exists at three levels in the domain: diverse users with different expectations and needs, different tools for users that help meet those needs, and different repositories of useful digital content on which users can draw, frequently with the intermediation of the tools. With this quite satisfying model in front of us, the Group formed two working groups to address the two interface layers in the model: one to think about the interface between repositories and tools, with an emphasis on what repositories should do to make their content optimally useful in such an environment; and one to explore the user experience through the medium of use cases. The latter effort led to an important observation about tools. The work of the two groups is discussed in the next two sections.

Case studies and the need for aggregation tools

In writing the use cases, this work group evolved a model of how resources are gathered and used in teaching. This model is a good deal more complex than the simple "find and incorporate" that is frequently assumed in much of the literature. It is based on three key

observations: that relevant digital resources will be distributed over many systems; that the process of using digital objects in teaching usually involves such tasks as arranging, editing, annotating, and describing; and that the results of this work may be used in multiple environments and/or saved for later reuse. The Group defined a general model workflow:

Gather:

DISCOVER = identify content sources
SEARCH = find content within sources
COLLECT = bookmark/link within each content source or within tool, probably using set formats or templates for types of learning objects or aggregations of content
IMPORT = into tool or managed environment, bring or point to content itself, or metadata about content
SAVE = prior to publishing, make a copy for the desktop, external or non-personal workspace that is managed for collaboration or sharing
FIND SIMILAR = identify like items, per the Amazon.com model

Create:

DESCRIBE RESOURCE = annotate, interpret, and write about content before publishing
ORGANIZE = order, sequence, transform content to create learning object
ASSOCIATE = declare link between content or learning object and course, project group or learning objective
MODIFY = change, edit, annotate content or learning object for re-use after initial publishing. Differentiated from **Organize** in that this function may trigger other services to selected community members such as Alerts or Notification related to allowable permissions or conditions to re-use

Share:

EXPORT = transfer content to other formats and/or tools, e.g., PowerPoint, METS. Differs from **Save** by its facilitation of supported format, output, packaging of content or learning object for specific display, rendering, use, storage environments
PUBLISH = make formally available to learning environment with implications for declaration / agreements related to rights for re-use, short and long term storage and archiving services, and expectations for content transformation services
ARCHIVE = establish agreements regarding short or long term storage, preservation, and delivery services.

The Working Group created three specific use cases of the use of digital library content for teaching (the full report of the Group is in *Appendix 4*). Taken together the three use cases illustrate this model and some of its ramifications. The first case describes the process of gathering readings for a humanities course. It involves searching in multiple sources (a library catalog, a digital library repository, and abstracting/indexing databases), and the integration of resources from multiple repositories (including the use of physical as well as digital resources). The second describes the use of a tool to support an instructor's work in aggregating images for use in a class. The tool provides an interface for searching for materials, supports the local aggregation of chosen materials, along with functions of arrangement and annotation, and provides options to output the aggregations in a variety of formats for different purposes. The third use case represents the use of a tool collecting resources that is embedded within a course management system, but that provides the means to search digital content repositories of many kinds including the subscription services managed by libraries.

The work involved in identifying where to look for resources, dealing with multiple system interfaces and varied search functionality, and incorporating heterogeneous metadata and objects into a local environment is enormously complex and rather daunting. The utility of a tool to simplify those tasks is obvious. Many sources, many interfaces, many digital formats are a given in our rich digital environment. Mitigating that complexity and diversity will certainly encourage and enable more instructors to make use of existing resources in their teaching.

A development related to the idea of an aggregation tool for instructors is the growing use of "metasearch" engines in libraries. These engines allow the simultaneous searching of multiple sources, with the engine masking the variations of interface and indexing across the various target systems, and homogenizing the metadata returned as a result of a search. This ability to mask the heterogeneity of many distributed information systems is an obvious part of any aggregation tool. It is worth noting as well the role "knowledge-bases" play in the metasearch environment: databases that contain information needed by the engines to find and use a variety of target systems. There is a good deal of work required to implement metasearch engines, work that will be common as well to aggregation tools. Such tools need to be configured to use appropriate target systems (i.e., configured for local needs and business arrangements), and the knowledge-bases need to be kept up-to-date as target systems change and evolve. An obvious question is how the work related to these distinct but related applications can be combined.

Predecessors of the aggregation systems discussed here are the citation-manager products such as EndNote and ProCite used by many scholars, which support searching, aggregation, homogenization, and flexible output of metadata from a rich variety of sources. Examples of aggregation tools proper are now being created, most notably the Scholar's Box system described in Use Case #2. We expect more such systems to be created in the near term. The Scholar's Box is a stand-alone system. It is easy to imagine similar tools being incorporated directly into course management systems, so that instructors have a unified environment in which to assemble all of the tools and resources needed to support a course.

Considerations for Repositories

As noted above, the universe of digital resources relevant to education is large, growing, and highly diverse. From the vantage point of a developer or operator of a course management system, enabling the use of such resources in local systems will inevitably be daunting. As the Group discussed the challenge and complexity of such integration, it became clear that there were a number of steps that the operators of repositories of content could take that would reduce the difficulties of locating and reusing their content. A work group was formed to analyze in detail what services and practices repository owners should consider when designing their offerings, and to create a checklist for repositories that includes specific standards or best practice recommendations when appropriate. The full report of the Group, including the checklist, is in *Appendix 2*, and a summary version of the checklist has been prepared by Kerry Blinco (IMS Australia).

Figure 2 below shows the relationship of systems and digital objects that the work group addresses. An important element of this diagram is the role of what the work group called gateway systems. These are systems that provide aggregation and discovery services for objects in distributed repositories. Examples of such gateways are abstracting and indexing databases such as Pubmed or Inspec, union catalogs such as OCLC or Melvyl, directories such as MERLOT, and even such search services such as Google. The checklist functions apply as much to these systems as to digital object repositories.

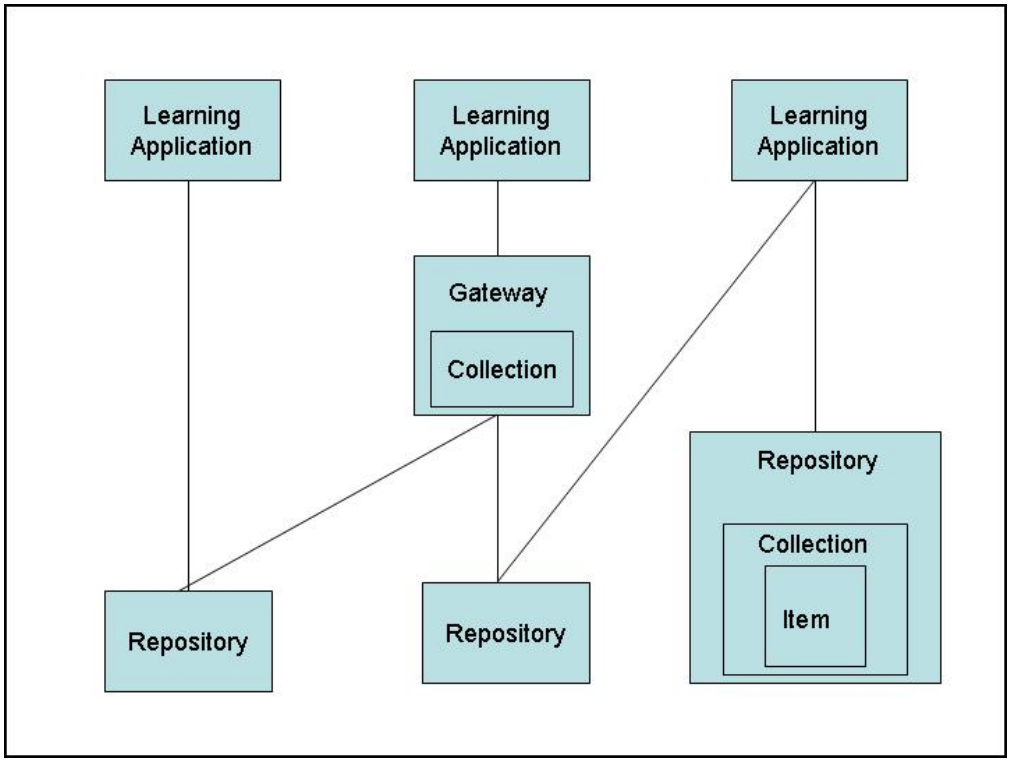

Figure 2. **Learning applications, Gateways, and Repositories**

The work group identified four key types of services relevant to the discovery and reuse of digital resources:

Finding Content. Issues include:

- what types of searching and browsing are supported, and whether complete inventories of content areas are available
- whether standard or well-documented descriptive, structural, technical, and administrative metadata are supported
- whether metadata can be exported for reuse in learning systems
- whether standard or well-documented protocols are supported for searching
- whether metadata is available for harvesting and inclusion in external discovery systems.

Collecting Content. Issues include:

- creation of stable identifiers for content, allowing it to be referenced unambiguously by outside systems
- creation of persistent resolvable identifiers for the location of content which will continue to reference content independent of changes in repositories
- support for standard citation formats, and the export of citations for exchange and reuse.

Accessing Content. Issues include:

- transfer of digital materials to local environments for manipulation and display
- ability of users to specify the format of materials or to download subsets of larger objects
- applications for viewing, utilizing and repurposing objects.

Documentation. Issues include:

- documentation of such critical policies as rights and use, privacy, and security
- documentation of all metadata conventions (vocabularies, subject classification, etc.)
- description of repositories in relevant registries, directories and gateways, so that users know of their existence and what content domains they cover
- description in relevant registries and directories of the technical and policy profiles of the repository (protocols supported, metadata standards supported, access policies) so that applications can interoperate with repositories appropriately.

In addition, the Group identified two general areas of design important for interoperation:

Accessibility. Issues include:

- does interaction with the repository require the use of proprietary protocols?
- does the user interface meet recognized accessibility guidelines and legislation?
- does the repository support standard character encodings?

Access Control. Issues include:

- does the repository interface to and support standard or conventional external authentication mechanisms?
- does the repository implement granular authorization rules?
- does the repository interface to standard or conventional external authorization mechanisms?
- does the repository interface support anonymous interaction for discovery?
- is authentication delayed until the point of need?

The overall thrust of the checklist is that repositories and related information systems should make themselves known to operators of learning applications in expected ways, should follow standards and best practices in terms of access, search, metadata practices, and download support, and should document their systems and policies so that others can configure their systems appropriately to interoperate with them. Taken together, these steps should significantly ease the task of integrating information systems into the learning environment.

In order to both test the checklist itself, and to get some feel for how the current environment of information systems relates to these criteria, a number of existing repositories in the digital library environment were asked to measure themselves against the list. *Appendix 3* includes the responses from six repositories: ARTStor, the California Digital Library, D-Space, Fedora, Harvard University, and JSTOR. These responses cannot simply be taken at face value for several reasons: they reveal different interpretations of some criteria; some of the responses are for software platforms and many of the criteria are specific to an implementation and some implementations of the software might comply and others; and the interpretations of functions that are "planned" obviously also varies noticeably. Nonetheless, there are a number of interesting observations one can make from the collective responses:

- Even among this population of players, who are likely to be aware of and sensitive to learning applications, no repository comes close to satisfying all of the listed criteria.

- These systems and services are evolving rapidly, and there are plans to implement a number of the criteria that not currently supported.

- One senses that there are a number of areas where, if there were accepted standards or practices that would enhance the usefulness of the repositories for learning applications, developers would be willing to add features to repository applications and that repository managers would be very willing to enhance their services.

- There are some obvious areas where the development of supporting infrastructure and community best practice is required. Preferred metadata formats and content (particularly for administrative data such as rights, and structural metadata for complex objects), a shared understanding of persistent identifiers, registries where services providing digital learning resources can make themselves known and record technical and policy profiles, and the role of and support for software agents to help deal with the large and complex information environment stand out as areas for development.

- Responses from the two institutions with large digital library environments (the California Digital Library and Harvard University) show that such environments are not homogeneous. Both responses repeatedly responded "in some cases", reflecting the variations across their systems. The landscape is complex even within single institutions.

Overall, these responses demonstrate the need for greater awareness of the issues of integration with learning environments, and for more active engagement between the digital library and course management communities.

These the guidelines for repositories and the aggregation tools discussed above are closely related issues. By supporting standards and community conventions and best practices, repositories and related systems can significantly simplify the task of building and maintaining aggregation tools that work across a large environment. The more target systems support standard services and document themselves, the larger the number of targets will be that can be practically supported by such tools.

The need for demonstration projects

The need for improved interoperation between learning systems and digital library systems has been much discussed, but we have today few working examples of such cooperation. As long as these discussions remain theoretical, neither the developers of instructional support systems nor the developers of digital library systems are likely to spend the resources required to support interoperation. We are at a point where some convincing demonstration projects are badly needed. The purpose of such projects include:

- demonstrating the utility of interoperation in the real world. The best argument for supporting interoperation will be instructors who use and care about the functions.

- testing the hypotheses about what functions matter. We need real world experience to see what is actually needed by instructors.

- providing experience with modes of interoperation. While the growing experience with metasearching is beginning to reveal what works and what doesn't in this sort of interoperation, and where we will need additional conventions, standards, and business models, the sort of aggregation tools imagined here will involve many issues beyond simple search, and only experience will show what is needed to support richer interoperation.

- providing a basis for projecting the resources required to implement and support wider interoperation. All the players in this environment are busy and stretched for resources. Interoperation represents a potentially large drain on resources, and systems designers and operators need to understand both the benefits **and** the costs of supporting it.

Our Group is convinced that there is now an adequate base of installed course management systems and of repositories of important educational content to mount meaningful demonstration projects in this domain, and we **strongly** encourage the Mellon Foundation to consider an initiative in this area. We believe that a variety of projects, involving different course management systems as well as a variety of content repositories, are needed. Content sources should include both commercial services (e-journal and e-book suppliers, art image collections, etc.) and university-based systems (institutional repositories, digital libraries, etc.).

Pages 3 – 7 of *Appendix 4* provide a framework created by the Use Case Working Group for considering an appropriate range of demonstration projects.

Other next steps

Discussions touched on a number of other efforts we believe will help further progress in this domain:

- *Use of digital library repositories to support reusable course content*. This Group looked at the interoperation of digital libraries and course management systems from one perspective: the inclusion of content from digital libraries in course environments. There is another potential area of interoperation that has been discussed repeatedly: the storing of materials created in a course context in digital library infrastructures for subsequent discovery and reuse. Such interoperation will involve an entirely different set of issues than those we considered, and we believe that a parallel effort to explore these issues would be beneficial.

- *Communication across domains and stakeholders.* We are struck by how few opportunities there are for digital library and course management developers and commercial information providers to talk systematically about areas of

intersection. We found that the various communities did not have a shared understanding of the larger environment, and that we had a great deal to learn from each other's world views. It is not easy to identify how to hold such larger discussions, but we believe an effort in this direction would provide significant pay-back.

- *The need for proselytizing*. Many information providers have little or no understanding of the role of course management systems, nor any appreciation of why making their content easily discoverable and reusable in a course context might matter. There is a need for librarians and course system operators to reach out particularly to the commercial information providers to begin to educate them about the growing role of such systems in higher education.