Learning Repositories Summit: Initial Research Summary

The Academic ADL Co-Lab
with support from
The William and Flora Hewlett Foundation

by Colin Holden
Academic ADL Co-Lab

Introduction

Information technology has entered the world of learning and it is here to stay. From the use of computing technologies in the classroom to the creation and delivery of entire online courses, new technologies are changing the ways in which we think about and practice education. With the development of the Internet has come the ability to easily deliver educational materials in electronic form to anyone, anywhere, at anytime. With developments in educational technology comes the promise that educational resources in electronic formats can change the ways in which we teach and learn.

But technologies do not transform societies or habits of practice all by themselves. The resources that develop from them must be easily located and retrieved, and these resources must be well selected to meet the needs of those persons to whom they are delivered. And those persons must know how to use the resources, and understand why it is in their best interests to use them at all.

This is why repositories, systems for the storage, location and retrieval of electronic content, are so essential to the further integration of information technologies and learning. They are the potential agents for breaking through many barriers to the use of new learning technologies.

While people are interested in repositories, and recognize their essential role in delivering educational content, few can agree on their precise definition. The search for an exact definition of the term "repository" has produced many potential candidates, some descriptive and some prescriptive. While theorists have debated definitions, developers and administrators have created networks that allow learning materials to be located and retrieved. In creating these networks, which they sometimes label "repositories," these individuals often make do in a world shaped more by financial and institutional constraints than theoretical concerns.

Whatever the exact definition of a repository is, all agree that **learning repositories can and should provide access to the increasing supply of digital educational content**.

There is much hope for the future of learning repositories. Tools are being developed, making available to educators and learners a broad range of learning materials. These learning materials are rich and varied both in content and form. A wide range of learning experiences is open to educators and learners.

This report draws upon research investigating projects that provide access to the increasing supply of digital educational content. It draws from this research conclusions about what has been done and needs to be done to permit greater access to high quality educational materials, and to encourage the use of those materials. Whether the projects already in place are technically repository projects themselves could be endlessly debated. Whether they should be investigated in order to inform future repository projects cannot.

It is hoped that this research and its conclusions will help us to better answer the following questions:

- How can learning repository leaders ensure the quality of the educational content that they provide access to?
- How can we ensure the quality and the consistency of metadata records, and encourage the use of metadata to better serve the needs of implementers?
- Should learning repository leaders create a distributed network of learning materials?
    - If so, what are the present barriers to the creation of this network?
    - If not, what kinds of access to learning materials should we be working to create?

These are questions of interest for all learning repository projects, independent of specific differences such as their content, audience, or tools.

I invite the reader to keep these questions in mind as they read this all too brief survey of what I shall call learning repositories. For the purpose of this paper, the definition of the term "learning repository" will be functional. **A resource is a learning repository if it is created in order to provide access to digital educational materials and if the nature of its content or metadata reflects an interest in those materials being used in an educational context**. This definition is a reflection of the current lack of agreement concerning the definition of learning repositories; it is not meant to discourage the creation of more precise definitions in the future.

The purpose of this paper, and of the Learning Repository Summit, is to encourage an awareness of the tremendous number of issues, and opportunities, that learning repository projects are facing. A learning repository project must take on many responsibilities, among them:

- To adequately describe educational materials
- To allow the location and retrieval of those materials
- For many repository projects, to encourage the use of those materials

Indeed, repository projects are involved in every stage of the delivery of learning materials and the encouragement of e-learning, except for the creation of the materials themselves. And some repository projects are involved in that as well. Unfortunately this work cannot be a comprehensive review of all the issues of concern to a learning repository project. The scope of this summary paper does not permit a full discussion of some subjects such as digital rights management and the long term sustainability of repository projects.

Repositories projects take advantage of computer technologies in two ways. First, as storage and retrieval systems, repositories can provide access to a large quantity of learning materials, creating records for these materials, applying metadata to them, and storing those records within a platform that allows them to be searched and retrieved.

Second, projects can manage their content in ways that allow that the content to exploit the electronic environment and alters the teaching and/or the learning process.

The intentions of learning repository initiatives are not solely the storage of quality digital materials. **Repositories are created in the hope that a community of practice will arise around them and that this community of practice will use the materials made available to it and, in many cases, improve the collection of those same materials**. Any successful repository project will have to be successful at creating this community. The technical and administrative tasks of repository management within the context of education as a whole are what make "learning repositories," into a subset of "content repositories," with roles and challenges specific to them.

This overview is not intended to be a comprehensive survey of all repository projects. A longer version of this paper, that will better survey the existing resources, will follow the Learning Repositories Summit (October 7-8, Madison, WI).

Learning Repositories: A Tentative Taxonomy

The variety of learning repositories is daunting for any researcher intending to make general statements about them all. Indeed any statements that apply to all run the risk of being helpful to none.

However, it is possible to describe a limited number of attributes that most repositories have but which vary repository to repository. Some of these attributes arise from the techniques that repositories use to store electronic learning materials and to locate them. Examples are:

- The structure of the repository itself. Whether the resources or the metadata describing them are on a single server, or distributed over many servers.
- The application of metadata. Whether metadata is applied at all, what kind of electronic record it is contained in (i.e. XML, HTML, Database), and who creates the content of this record.
- Metadata standards adopted. Whether the repository's metadata schema uses or builds on an existing metadata standard such as Dublin Core or IEEE's LOM (Learning Objects Metadata).
- Interoperability. Whether the electronic record containing metadata is readable by an off-site search engine, and the metadata contained follows interoperability standards.

Other attributes that vary repository to repository arise from the various ways in which repositories define and practice their relationship with a wider community of educators and/or learners. Examples of these are:

- The repository's intended users. These could be educators who are choosing course materials for their own classes, or learners who will access learning materials directly through the repository, etc.

- The effect of the repository's contents on pedagogical practice. Whether the contents of the repository reflect a specific pedagogical theory or encourage a specific pedagogical practice. This is related to the consistency of repository contents in terms of quality, granularity, etc.
- The capacity of the repository staff to create a community that is aware of its materials and is familiar with how to use them.
- The capacity of the repository staff to create or be part of a community that creates learning materials and is willing to deposit them within the repository.

If we think of each repository as existing in some relation to these eight issues we will have simplified somewhat our picture of the repository projects currently in existence.

A review of the materials and resources currently available online shows that many repository projects excel in some of these domains but few in all. Some have excellent systems of outreach and peer review but mixed collections of materials. Others have extremely good content but are not well known. Some hold fast to their metadata schema, applying it well and consistently, but lack a large collection of materials. Yet for all of these divergent directions, the core purposes of these projects, providing access to learning materials, remains the same.

The various differences among projects may be obstacles to creating unified networks that bring together materials from various collections in ways that serve the needs of users. If repositories are to share collections it is important that we are cognizant of what institutional and technical structures encourage or discourage the unification of resources.

How Can Learning Repository Leaders Ensure the Quality of the Educational Content that they Provide Access to?

To encourage the use of repositories and thus of online educational content, repository projects should provide users with "quality" materials. However, the definition of "quality" is contextual, a function of the needs of specific learners and specific educators in specific and sometimes unpredictable contexts. An involved and interactive lesson on pendulums might be a "quality" resource for a learner studying physics, but for a teacher who simply needs a visual aid it will contain extraneous and distracting features.

To deliver quality materials a learning repository should

- Predict the needs of its users
- Provide users with the means to describe their own needs when searching the repository
- Deliver content that meets those needs

To implement these tasks repository projects must adopt institutional structures capable of meeting user needs.

Almost all learning repositories either receive their materials through submissions from users or plan to receive submissions in the future. This situation places a lot of responsibility for content development on the hands of persons who the repository project does not have control over. Thus the project must create its desired community of practice within a population that is hard to control.

Repository projects have devised many means for coping with this problem, and these means will dominate this paper. But we must note that some repository projects do all of their collection development themselves; selecting, reviewing, and cataloging materials under a single roof. The Eisenhower National Clearinghouse for Mathematics and Science Education, as well as LearningLanguages.net, for example, both have staffs of subject matter experts who locate and select materials for their collections. The size of both project's collections suggests that the cost of the strategy is not prohibitive.

However, as the majority of projects do take and even depend upon user submissions for their collection development, we must ask whether this strategy can work in concert with a desire for collections made up of quality materials.

- **Predicting user needs**

Learning repositories have tried to predict the needs of its users in a number of ways. They include:

- Allowing users to submit, rate, and comment on resources, thus allowing members of the user community to specify multiple definitions of "quality" themselves.
- Delivering content only after it has been reviewed by repository staff, or reviewing some amount of the repository content.
- Delivering content to a pre-defined population (defined by subject matter or educational level) whose expectations can be more easily predicted and catered to.

Many users have taken advantage of sites that allow them to rate and comment on resources, techniques that have been used by commercial Web sites. This allows sites to provide some level of direction to quality resources while reducing the labor demand on on-site staff. For repositories that provide access to many thousands of resources the cost of reviewing all materials might be prohibitive. MERLOT augments user reviews by providing "peer-reviews" performed by onsite staff who are specialists in the subject matter of the materials they review.

Other repository projects target specific professions or fields. Repositories have been established that deliver materials to educators and learners in such specific fields as the Earth Sciences, Computer Science, and Engineering. By narrowing the range of a repository's content, users can be assured that as long as they are members of that repository's target audience, they are more likely to locate the kinds of resources they desire.

- **Describing and finding materials**

The most powerful tool currently available for describing and locating digital materials is metadata and a thorough metadata schema. By creating records and making them available for searching, the creator of metadata allows a resource to be discovered by a user who might not know in advance the specific resource they require. By creating metadata records that describe multiple attributes of the resource referred to, the creator of metadata allows a resource to be discovered using multiple searching techniques.

A sufficiently nuanced metadata schema should permit users to locate materials using a variety of terms or phrases as search terms. The contents of each of a metadata record's fields describe various attributes of a resource such as its title, its creator or creators, or its subject. Ideally, a metadata schema would take into account the needs of its presumed users. LOM, for example, includes fields of particular interest to educators or learners such as the intended age range of a resource's users and the resource's interactivity level.
.
To aid in the location of desired materials, metadata must meet the needs of searchers. However, repository metadata may meet many needs. Creators of metadata records may also be concerned whether:

- The metadata record is compliant with standards that will permit interoperability with other systems and thus allow records to be searched from multiple locations.
- The metadata record allows resources to be collated or classified with other similar resources through the use of consistent or controlled vocabularies.
- The metadata record accurately describes the material it refers to once it has been located.

These priorities do not always lead to the creation of easily discoverable resources. The use of controlled vocabularies, for example, is very useful for consistently classifying a growing quantity of materials, and ensuring that when a controlled term is used the found materials will share common attributes. But when users are unaware of a controlled vocabulary's terms or meanings, the vocabulary will be unable to serve their needs. A look at Library of Congress subject headings, with such terms as "Cookery" for what would more commonly be called cooking, shows that the terms with which vocabularies are controlled and thus subjects classified are not always the terms with which users will search for records corresponding to those same subjects.

We can describe this issue as a conflict between a desire for precision in resource description on the one hand, and the desire that users can search successfully for resources using a variety of search terms. Some basic tools exist to allow metadata records to be friendly to multiple searchers while still allowing for precision in resource location and description when they are required. The description field, and the keyword field, used in Dublin Core and LOM, allow for the use of natural language phrases and terms, as well as vocabularies specific to a field or community. Use of these fields grants the creator of the metadata record room to increase the chances that a record will be found while allowing the contents of other metadata fields to be precisely defined.

Controlled vocabularies, when used alongside thesauri, can also allow for both precision in classification and location of materials through a broad range of possible searches. However, while it is not difficult to imagine thesauri functioning in the digital context, the costs of creating and implementing them are large. Some resources have been developed, however. The ERIC Processing and Reference Facility has an online thesaurus of educational subjects, for example. EdNA also uses thesauri terms to encourage good resource discovery results. EdNA has reported that this "has its challenges – each sector has chosen a preferred thesaurus."

- **Delivering desired content**

A collection is as strong as the resources it can deliver, and a repository's collection development strategy will have an impact on what resources it will contain. The quality of a repository's collection, that is the likelihood that its contents will be of value to users, depends on the decisions that repository administrators make as far as:

- What qualifies a user to submit materials, and motivations for submission.
- What review process exists for submitted materials.
- What criteria are used to determine inclusion of materials within the repository collection.

Most repositories require that a user submitting resources be logged in to the repository interface. Logging in requires that users have created a user profile providing some basic information about themselves and, when required, specified that they are part of a member institution permitted to submit resources to the repository.

- **Resource Submission**

Permitting a body of volunteers, even of logged on members, to submit records for resources has its advantages, as is evidenced by the sizes of many repositories. However there are issues that must be faced by any repository project using this strategy.

- It is sometimes unclear or very doubtful whether the submitter of a record for a resource is the creator of that resource or instead simply its discoverer. This opens the possibilities that the same resource, or multiple versions of the same resource, could be submitted by multiple persons. Some techniques should be developed for both detecting duplicates and multiple versions.
- A body of volunteers might not be familiar with the priorities and philosophies that underlie a repository project's mission. Among other issues, it might be difficult to enforce any definition of "Learning Object," or maintain a consistent level of granularity throughout the collection, nor might they create or submit materials that meet repository project requirements for accessibility
- Volunteers from the educational community, most of whom are not developers, do not always produce or submit materials that take full advantage of the digital environment. If it is in the interest of the repository project to encourage the use

of interactive electronic materials that do not have real-world analogs, means must be developed to aid users in developing new kinds of materials, or appeal to a group of users that has the capacity to create interactive media.

- The legal and copyright status of electronic learning materials as intellectual property remains ambiguous. Creators of these learning materials might be discouraged from sharing them.

Duplication of materials within repositories is a difficult problem to solve. It is very easy for two people, especially volunteer submitters, to produce different records, even to assign different titles to resources, given the ambiguous structures of electronic documents. A small handful of repository projects have made available to submitters the AACR2 cataloging rules, which specify the techniques by which the title, the creator, and other attributes of a digital document will be specified, including the forms that names and titles will take. But AACR2 is complex, and it is unclear to what extent making the materials available online has encouraged widespread compliance with its rules.

Duplication prevention might also be assisted if electronic documents were assigned Uniform Resource Identifiers (URIs), Uniform Resource Names (URNs), or some other globally unique ID. It is unclear at present to what extent these identifiers would work on the Web as a whole, and how volunteer submitters of mixed technical backgrounds would be able to identify and use them when submitting sources.

Some techniques have been developed for encouraging consistency in granularity across collections of learning. One successful, though indirect, method to encourage consistent levels of granularity has been to collect specific kinds of software that deliver educational content or activities. Thus, collections of Java Applets or Flash files, for example, will contain materials with certain uniform attributes and relatively consistent levels of granularity established by the software.

One project, the Wisconsin Online Resource Center/Catalog of Objects has a permanent staff that takes ideas for learning objects from educators and then creates a Flash learning object according to the project's own specifications. This maintains consistency within the collection, though it does slow down the production of materials.

A survey of the contents of the largest repositories shows that many of the online materials currently available for teachers, both in learning repositories and elsewhere, have real-world analogs. Such materials include lesson plans that are made up entirely of text, tests or review sheets that must be printed out to be used in class, or information for teachers about instructional technique or pedagogy itself.

The Internet is a delivery system and it should come as no surprise that when educators are given access to it they will use it to exchange materials that they already exchange by other means and which they are already familiar with. Many educators are trained in traditional teaching techniques and thus lack sufficient drive to exploit potential that information technologies have to not only deliver content, but to run new kinds of interactive content that exploits the electronic environment.

The limited amount of available classroom technology in many classrooms, especially K-12 classrooms, limits the potential demand for, and the capacities of many educators, to create materials that exploit digital technologies. If the technology is not available, educators will demand materials that can be delivered to students through traditional pedagogical means.

Some repository projects have responded directly to this demand for traditional learning materials. The Gateway to Educational Materials, for example, collects lesson plans, projects, and other materials to be used in class.

Many repository projects have charted a course towards changing the ways in which learning technologies are used in the classroom, and have done so by collecting resources that meet specific criteria for either quality or educational style. The UK's National Learning Network, for example, provides users with a collection made up of interactive, browser based materials that can either be integrated into the classroom setting or accessed by learners directly. Another project, Connexions, contains materials whose content and metadata fit a very specific XML format that permits resources to be easily edited, linked together in sequences, and allows content to exist independently of any one template, background, or lesson design. CLOE has proposed a tentative system of incentives to create online learning materials of high quality. Users would participate in an online exchange of learning materials and continued participation in that exchange, and thus continued access to learning materials, would depend upon submission of materials that were accessed and used by others.

Means for dealing with concerns about digital copyright remain under discussion. One solution, adopted by the Connexions project, is to encourage users to use Creative Commons licenses for all materials submitted to their site. Creative Commons has developed machine readable metadata that can associate creative works with their public domain or license status.

- **Review and review criteria**

There are repository projects that review some or all of the materials submitted to them. The Learning Matrix, for example, has a staff that reviews submitted materials. Of course, increased project staff can both drive up costs and slow down the posting of materials into the repository.

Many sites have developed a review process for submitted materials, thus allowing them greater control over the contents of their collections. This situation allows for many benefits in the realm of control over a consistent and managed collection of materials.

We should not see user input as a potentially corrupting force for repository collections. Users respond to repository projects that accept their input, not only by allowing them to submit resources but also allowing them to comment on and rate resources, as well as create their own collections of recommended resources. MERLOT allows users to create

their own personal collections of favorite resources that other users can then browse through. MERLOT users can also list assignments or units of instruction that they have created around the materials referred to in the repository collection. This creates a community actively involved in finding ways to bring digital content into the educational environment.

To encourage the use of digital content in teaching and learning, a repository project would do well to encourage the development of a community that develops, uses, and improves upon digital learning materials. Making the repository site into a place for information exchange and community building among educators increases the likelihood that users will visit the site, and thus use its materials and submit to it.

- **Issues for consideration**

It is clear there is a great quantity of digital education content. But as we develop ways to collect these materials we must be concerned with:

- What is the role of the general public in developing a repository's collection? Does public input further the mission of all repository projects? When does it hinder it?
- How can we create metadata records that best meet user needs to identify and retrieve those records electronically through a variety of search techniques while allowing these records to be classified according to consistent rules?
- What balance can repository projects strike between allowing users to exchange learning materials they deem useful, and encouraging the use of materials that users might not be familiar or comfortable with? Are all projects even required to be concerned with the later?

How Can We Ensure the Quality and the Consistency of Metadata Records, and Encourage the Use of Metadata to Better Serve the Needs of Implementers?

When discussing metadata we must distinguish between the various roles that metadata plays:
Some uses of metadata and metadata standards include

- Describing materials in ways that aid users in identifying those that meet their needs.
- Providing access to those materials through multiple access points, and thus through multiple potential search strategies.
- Collating, or classifying, similar materials
- Permitting compliance with metadata standards allowing interoperability with multiple collections.

It is also important to note that the implementation of metadata can be described as being made up of three tasks

- The adoption of a metadata schema
- The creation of a workflow through which metadata records will be created
- The actual creation of the metadata records

Finally, it is important to specify why metadata must better serve the needs of "implementers," and what is meant by that term.

The "implementer" of metadata is the person or persons primarily responsible for the creation of metadata records. In a library setting this would be the cataloger. It is important that we think about the motivations of the implementer because their motivations may not always lead to the creation of records that meet the various needs of searchers, administrators, and those harvesting metadata.

These various and complex issues shall, for the purposes of this review, reduce down to these concerns

- Who should produce metadata records?
- What tools should be available for the creator of metadata?
- How can the creator of metadata records be encouraged to create records that meet standards and/or the needs of various users and administrators?

  - **Who should produce metadata records?**

The metadata that preceded digital technologies, library catalogs and the like, were produced by trained groups of professionals. Book catalogers do not discuss whether the library record of a given book should be created by that book's author or by those who donate the book to the library. The extent to which learning repositories currently depend on non-professionals for the creation of their metadata is unprecedented. We can view it to some extent as a giant social experiment, created by institutional and budgetary constraints, of the ability of the general public to create useful metadata.

Much like with content, the definition of "quality" within the realm of metadata records is difficult to specify. As the definition is contextual we can define a quality metadata record as a record that is useful in a number of different contexts, both with respect to the search strategies and terms that can be used to locate it. Here I shall briefly discuss some issues that we must confront if we are to grant general users the responsibilities of creating metadata that meet the various demands for quality metadata.

Many learning repositories allow creators or submitters of materials to fill the metadata fields within the records for the materials that they submit. This allows the repository to receive ready-made records for its materials, thus avoiding the logistical issues surrounding the creation of those same records by on-site staff. However, it is difficult to both allow the public to create metadata records and maintain a certain level of quality

and consistency within the metadata of the collection. Some collections that leave metadata submission open to the public suffer from blank fields and fields containing information of ambiguous quality.

A cursory investigation of sites that depend upon volunteers for the content of their metadata suggests that while many are comfortable with the "description" fields that many schemas make available, providing long and discursive descriptions of their resources, they are less willing to fill out other metadata fields with more specific kinds of information. Thus records are created which are descriptive, and also findable by someone using the same vocabulary as the submitter of the resource, but are not responsive to search techniques that enhance precision or are aimed at finding pre-defined classes of resources.

While many repository projects provide documents guiding and advising users in the creation of metadata and the maintenance of any controlled vocabularies the specific project might maintain, it is unclear to what extent the availability of this information actually affects user behavior. Techniques are required to create a culture of practice that creates metadata that meet the needs of both repository administrators and users.

It may be that repository metadata does not need to hold to specific practices in the creation of metadata, precise vocabularies, or subject classifications in order to be useful. To assess to what extent repository searchers need or take advantage of chances to make precise searches, searching within multiple metadata fields simultaneously, and making use of controlled vocabularies, more work must be done to study how learning materials are searched for.

- **Tools and Workflow**

Many repository projects have developed systems for the creation of metadata that involve multiple persons, each assigned distinct tasks. Some have allowed submitting users to assign values to some metadata fields, but assign the filling in of others (requiring what the administrators believe to be some level of expertise) to onsite staff. Others allow onsite staff to review, assess the accuracy of, and correct metadata provided by the resource submitter.

Many repository projects have developed interfaces for the creation of metadata. Thus, normally through a series of forms, users go through a step by step process in which they must fill out some metadata fields, and are constrained by drop down boxes to use controlled vocabularies, as long as those vocabularies are relatively small.

Such tools include GEMCat, developed by the Gateway to Educational Materials, cataloging interfaces for the Heal repository, and the resource submission page for MERLOT.

- **Vocabularies and metadata consistency**

As the quantity of learning materials increases it is becoming evident that attempts to categorize them should be aided by ontologies and controlled vocabularies. While some sites have their own limited schemas for defining and classifying the subjects for their materials, the growing collections of materials seems to reveal a need for systems that are as comprehensive as the subject classification schemas used in libraries and suited to the electronic environment.

When developing taxonomies, classifications, or ontologies for repositories it is important to discern what these tools are for. A taxonomy or ontology allows similar materials to be retrieved together through the use of a controlled search term, and it allows materials to be browsed by browsing a schema.

The creation of a classification schema for learning materials is a monumental task. The complexity of such systems as the Library of Congress Classifications and Subject Headings reveals how much work and institutional support can go into just one attempt to classify the subjects the define human knowledge.

There have been moves to creating taxonomies and ontologies of universal scope. LearnDirect has developed the Learn Direct Classification System (LDCS) which it uses to classify its own training and educational materials. The system is free to all in text format. But such schemas have yet to be widely adopted. There are, however, controlled vocabularies and taxonomies that have been adopted by various professional or academic groups. MedBiquitous, in its own metadata specifications, permits the use of the Mesh, SNOMED, and UMLS controlled vocabularies to describe medical subjects to facilitate searching for and retrieval of materials by persons in the medical profession.

Much of what allows these tools to work can be located within the institutional structures that surround and enforce a given standard. The subject classification adopted by the Library of Congress is not the prominent and frequently used system it is because it is the best system possible. Indeed, any librarian can point out its flaws and inconsistencies. The Library of Congress classifications and subject headings are widespread because of institutional structures that educate persons in their use and create a network of libraries all using the same system. We can learn from libraries to aid us in the cataloging and classification of materials. But the secret may not be to study the lists of subject headings and classifications they have developed, but to look at the institutional structures that transform those lists into real arrangements of real materials across various library collections.

- **Issues for Consideration**

- How can creators of metadata be encouraged to create metadata that meet the needs of other users and of administrators?

- How important is it that users be capable of performing precise searching? Does its importance justify the institutional and practical work that would need to be done to create a collection capable of responding to precision in searching?

Should Learning Repository Leaders Create a Network of Learning Materials?

o Distributed Repositories and Silos

The number of learning repositories is rising. Even as well-established hubs such as MERLOT provide access to educational materials, repositories continue to be developed with their own contents, and their own metadata. Educators and developers will continue to create their own repositories, and these hubs alone cannot provide access to all available learning materials.

In the face of the constant rate at which new repositories are developed, it is becoming more and more evident that if repositories are to provide access to the largest possible collection of learning materials, means must be developed by which the metadata of multiple repositories can be searched at once.

One of the most exciting developments in the realm of learning repositories, and repositories in general, has been the development of distributed repository architectures. A distributed repository consists of a network of physical repositories, which are capable of sharing metadata among themselves, and that together can function as a single repository. They allow users to search metadata stored within multiple physical repositories.

The distributed repository model stands in contrast to the more common centralized model, which has been described as the "silo." A silo is a centralized collection of metadata and resources, centralized either because it is located within a specific device or available only through a specific institution or organization. While the database that supports each "silo" may be distributed, a repository can be described as a silo if the repository can only be accessed through a single point of access, usually a single URL.

At present, nearly all repositories can be described as silos, providing access to resources, or the metadata describing those resources, through a single gateway or URL. However, using protocols for metadata harvesting, some repository projects have made their metadata accessible to those who are not accessing the repositories Web page.

Protocols for metadata harvesting require that servers provide metadata that harvesters (server processes), can collect. The Open Archives Initiative (OAI) provides the protocol most commonly used in metadata harvesting among learning repositories.

At present, repositories that contain harvestable metadata and harvest other metadata records include MERLOT, EdNA, ENC, and The Learning Matrix. CAREO, in its final form, will be a network of collections, each able to access each other's harvestable

metadata. NSDL also sponsors many sites, each of which is required to be OAI-PMH (Protocol for Metadata Harvesting) compliant.

Distributed repositories can arise from the interconnection of pre-existent collections. But there are plans to develop distributed repositories that do not link previously centralized collections, but create instead decentralized stores of metadata and content. Portals for Online Objects for Learning (POOL), a Canadian consortium project funded in part by the Canarie Learning Program, has developed a Peer-to-Peer structure for distributed metadata searching. It has facilitated the creation of various collections by making available SPLASH, a downloadable application for storing, searching and exchanging objects using the CanCore metadata schema. This tool permits metadata creation, storage of metadata and searching of the POOL network. SPLASH is a desktop client that communicates with other peers via the Peer-to-Peer POOL protocol.

It is possible to create distributed repositories, with multiple nodes where new metadata and new materials might be entered, the entire repository being searchable from any point within the system, or outside of it using the appropriate tools. At this point it would be wise to assess the benefits of these distributed systems and determine whether they are beneficial in any or all of the various domains in which repositories aid education.

- **Distributed Repositories, Distributed Institutions.**

The ability of an organization to act hinges on its institutional structure. Thus, the capacity of a repository to deliver content and meet the needs of its users depends first on the structure of the repository project itself. New technologies are allowing us to imagine new repository architectures, and it is important that we keep in mind that with these technological infrastructures will come new institutional structures, workflows, and distributions of power over repository content.

With the centralization of stored information comes a centralization of labor. If a repository project has a well-defined mission, and pursues it by delivering a specific kind of content, it is in the interest of the repository project staff to maintain firm control over the repository content and the metadata that describes it. It is easier for on-site staff to control these factors if the repository is centralized.

Distributed repositories move us closer to creating a large and comprehensive network of learning materials. But if we are not careful they may also remove many of the benefits that come from the centralization of energy and resources within a single hub. Distributed repositories make available and deliver more content, but there are less technical innovations in metadata creation, collection management, and outreach that repository projects have made which we do not want to loose sight of.

Creating distributed repositories raises the possibility that the contents of a distributed collection's metadata, both in terms of completeness and quality, as well as the completeness and quality of the content referred to, will vary with to the policies of the individual collections from which the metadata are harvested.

It is important to ask, as we develop more and more distributed repositories, how collections drawn from different institutional structures are to be combined. Should users and administrators have information about the collections that make up a distributed repository? Will they be able to use this information to search one collection but not another? Or can the various collections within a repository work in concert in collection management or metadata creation?

- **Distributed Repositories and Metadata**

One of the consequences of the drive to harvest metadata and encourage distributed repositories has been the development and promotion of communication protocols and metadata standards in the interests of promoting compliance across multiple repositories. This compliance of course, makes it easier for multiple repositories to share their collections with other repositories or standards compliant tools.

This has led to a change in the way some repository project leaders talk about metadata. Compliance with metadata standards is obligatory if distributed repositories are to function. Thus discussions of metadata can frequently turn into discussions concerning *compliance* with metadata standards. Yet as these standards are encouraged in the interests of interoperability we must not forget the other functions of metadata, to describe and locate resources, and to collate records.

The metadata standards with which we tie together the various collections of a distributed repository will either augment or limit the powers of users to describe and locate the resources they require. With the development of RSS feeds, which have no metadata requirements, the task of structuring metadata across collections becomes a more complex.

Some repositories have developed metadata schemas that facilitate both interoperability and the other uses of metadata. EdNA, which harvests metadata records from a number of Australian collections of learning materials, requires that harvested materials be compliant with Dublin Core or the EdNA Metadata Standard Version 1.1, which has been developed to meet the needs of educators and learners locating materials.

- **Issues for Consideration**

As we assess the powers of distributed repositories we are rightfully impressed by the quantity of materials they can potentially find within multiple collections. But it is worth asking:

- Does a distributed repository require similar institutional structures or habits of practice within each of its collections? If not, why not? And if so, how will uniformity among habits or practices be encouraged or enforced?
- Are the metadata schemas with which we connect our collections useful to users who are describing and locating resources?

Concluding Remarks

Educators and learners demand a wide variety of digital content, and as digital technologies further establish themselves in learning environments, learning repository project leaders must decide how they will respond to a demand for content that is growing both in size and diversity.

The next few years will likely see the creation of distributed networks of learning materials. But the kinds of networks these will be and the needs they will serve will be functions of the technical and institutional structures that make them up. There is a difference between exchanging information and creating lasting catalogs of resources. The balance that is struck between free submission of resources and metadata on the one hand, and control over the quality of both on the other, will determine to how the repositories of the future will situate themselves between the extremes of the library catalog on the one hand, and the searchable weblog on the other.

Beyond the question of creating technical networks of learning repositories, it is important that learning repository project leaders create an intellectual community that exchanges information, experiences, and expertise. This is essential for the development of learning repositories as every project cannot be expert at every task and challenge it is presented with. As the philosopher of science Philip Kitcher writes, a "virtual consensus" among thinkers does not require that those thinkers "clutter their heads with the details of other people's subspecialties. When they need information, they go to people who are regarded as authoritative with respect to the issue involved." Learning repositories will develop if we map out an intellectual community that can function as a whole. Thus expert knowledge will be constituted not of the contents of each repository project leader's head, but of everything a thinker can discover by accessing and following the judgments of a chain of authorities. With conferences like the Learning Repositories Summit we can hopefully move closer to creating connections that allow us to create a community in which specialized expertise within specific projects is recognized and drawn upon by the entire community and a community of practice is created among the leaders themselves.

One of the most important tasks for any repository project, one that supersedes all others mentioned here, is survival. Repositories must have a constant base of support. Within the United States, there are still many repositories that are grant funded, and are not guaranteed continued funding in the future. In Europe and Canada repositories are more likely to be associated with government and academic institutions, that are presumably more likely to fund repository projects on a long-term basis.

If we cannot sustain the repositories we create we will retard our progress towards providing access to digital materials. If collections on which individuals have worked hard are forgotten or neglected, as some already seem to be, we will end up recreating, rediscovering and re-describing materials many times over.