Connecting and Extending Peer-to-Peer Networks

A Penn State Proposal to the Andrew W. Mellon Foundation

September 2003

Abstract

With support from the Mellon Foundation, the Visual User Image Study (VIUS)[1] discovered how digital images are used at Penn State for teaching, research, and outreach. Through extensive contact with stakeholders, the VIUS team identified criteria that point to the need for more flexible, user-controlled tools and services for discovery, organization, management, and sharing of personal and public images. Many of the features that stakeholders identified suggested a much different model of collaboration is needed—one that places much more control in the hands of the end user than currently available in large central repositories and that peer-to-peer (P2P) technologies offer many of these features. The VIUS team determined that peer-to-peer technologies offer unique opportunities to enable greater access and use of personal and publicly available image collections and that a prototype was warranted because the vast majority of digital images reside in private collections and that access to and integration of public image collections is problematic. The LionShare proposal emerges from the lessons learned from VIUS.

LionShare: Connecting and Extending Peer-to-Peer Networks extends the definition of a learning resource beyond digital images to include a wide variety of digital resources including images, audio, video, simulations, text, documents, research papers, Web resources, and other learning resources and activities. The result will be the development of a public release of the LionShare software that will include: 1) an individual LionShare Peer—for personal file management, search/retrieval, and collaboration; 2) a LionShare PeerServer—for aggregating resources for small and large groups, 3) networking to support P2P collaboration, and 4) software components to enable integration with central and distributed learning object repositories. While VIUS was restricted to the Penn State environment, LionShare will leverage the work of the Internet2 community by using Shibboleth to create a much broader community of authenticated, authorized users.

The primary research question that this proposal wishes to address is:

> *How can P2P technologies be designed to best promote collaboration for teaching, learning, research, and outreach in the Higher Education environment?*

The project will have four major elements.

1.) *LionShare P2P Network Development* will focus on continued development of the LionShare P2P architecture;
2.) *Federating P2P Networks* will address the need for a middleware infrastructure to allow sharing of resources among universities;
3.) *Connecting Learning Object Repositories* aims to address the need to create bridges between multiple types of repositories, both central and distributed, for the retrieval of appropriate resources by users of P2P networks; and
4.) *Project Assessment* will assess how successfully the project meets the needs of users.

A major part of the LionShare project is to build a sustainable P2P technology infrastructure through collaboration within a community interested in the technology agenda and the broader benefits that participation brings. All technologies will be developed under the Open Source GNU GPL license so technology developments can be highly leveraged in years to come for other purposes. Cooperating groups committed to this project, thus far, include the Internet2 P2P and Middleware Initiative Working Groups, the eduSource Canada community, the Canadian Learning Object Repositories Networks (LORnet) group—now being formed, and the Open Knowledge Initiative. These collaborative development efforts provide a significant opportunity to advance the state-of-the-art in the development advanced P2P networks, multi-institution access control, and the interoperability of a multi-national learning object repository.

Table of Contents

LionShare:
Connecting and Extending P2P Networks

Section 1 – Project Overview

VIUS and LionShare

Background

The Visual User Image Study (VIUS), a 26-month Mellon funded project ending September 2003 house in the Penn State University Libraries, is discovering how digital images are used for teaching, research, and outreach. Major components of the project include extensive interviews with stakeholders (students, faculty, and collections managers), surveys, focus groups, transaction logs, and protocol analysis to determine how individuals and groups are currently using images, where they find such images, and what they would like to do with digital images now and in the future.

Through this extensive contact with stakeholders, the VIUS team identified a number of criteria that point to the need for more flexible, user-controlled tools and services for discovery, organization, management, and sharing of personal and public images. Many of the kinds of features that stakeholders identified suggested a much different model of collaboration is needed—one that places much more control in the hands of the end user than currently available in large central repositories and that peer-to-peer (P2P) technologies offered many of these features. The VIUS team determined that P2P technologies offer unique opportunities to enable greater access and use of personal and publicly available image collections and that a prototype is warranted. The study also revealed the need for a more secure approach to ensure long-term viability and to protect the interest of personal collections. The study further determined that the vast majority of digital images reside in private collections and that access to and integration of public image collections is problematic.

The VIUS LionShare prototype evolved from these broad assessment activities as a potential solution to these problems. It is this prototypical work, as a result of the VIUS LionShare project, that this proposal, *LionShare: Connecting and Extending Peer-to-Peer Networks*, seeks to extend with further development activities. It also seeks to extend the definition of a learning resource beyond digital images to include a wide variety of digital resources including images, audio, video, simulations, text, documents, research papers, Web resources, and other learning resources and activities. The result will be the development of a public release of the LionShare software that will include: 1) an individual LionShare Peer—for personal file management, search/retrieval, and collaboration; 2) a LionShare PeerServer—for aggregating and securely sharing resources for small and large groups, 3) networking to support P2P collaboration, and 4) software components to enable integration with central and distributed learning object repositories.

A Mellon Proposal
LionShare: Connecting and Extending P2P Networks

Problem Statement

Many instructors, scholars, researchers, and librarians across higher education institutions (HEIs) have "hidden" repositories of digital content (images, audio, video, research papers, learning resources, learning activities, etc.) used for teaching, research, and outreach stored on their networks or even individual hard drives. This content is "hidden" in the sense that other potential users at their own and other institutions have no way to discover these resources. Additionally, if they could find these resources, it is not clear whether they could use the materials or how they would get permission to do so. This last problem highlights the reason that it is not prudent to simply migrate that content to a public Web site if an individual or organization wishes to maintain some ownership rights in the materials. In short, there is a distribution problem for the digital content, which typically has several possible solutions: central repository, distributed repositories, peer-to-peer distribution networks, grid networks, and so forth. Furthermore, because of the wide variety of distribution solutions, resource discovery is further complicated by the fact

that there are no federated search strategies that allow a single search query to reach all available repositories.

Why P2P Networks in Academia?

In 2001, Andy Oram gave a couple of key presentations, one at the O'Reilly Peer-to-Peer and Web Services Conference and the other a keynote speech at the Internet2 Fall Members Meeting, highlighting the rationale for research activities related to P2P technologies within academia. Oram states:

> *"Academic environments are ideal for experimenting with peer-to-peer and benefiting from peer-to-peer. You have an open attitude toward information, well-educated staff who can adapt to new tools, a variety of projects that require information exchange, and a willingness to expend time and effort in order to save money…The Internet2 project, in particular, overcomes many barriers that are holding back the deployment of peer-to-peer products in current corporate environments. Internet2 is a good test bed for basic research that can benefit peer-to-peer." Andy Oram. Peer-to-Peer for Academia. 2001.*

These presentations and Oram's book entitled, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies (2001)*, stimulated thinking and discussion in academic communities about the potential for P2P networks as valuable resources for collaboration and sharing. Much of that discussion has led to groundbreaking research into the efficacy of P2P networks and grid computing (also a P2P technology). The VIUS project at Penn State and the Canadian eduSource project further validate the interest and potential in nascent distributed networking approaches for file sharing and collaboration in their assessment and development activities. Thus, academic P2P network for personal, small group, and public learning object repositories offer attractive solutions for the following identified reasons:

- Content remains close to the creator
- Individuals can manage and control access to personal libraries of learning object (LO) — those they create and collect
- P2P solutions require little technical support but give access to a wide variety of LO across informal networks
- P2P users can create formal and informal sub-networks to enhance search and filter content
- Canadian LO research demonstrates that P2P networks can be federated with more formal centralized repositories
- P2P networks are less brittle and have the capacity to accommodate peaks demands since resources can be distributed and accessed simultaneously from multiple sites at the same time.

Primary Research Questions

The primary research question that this proposal wishes to address is:

> *How can P2P technologies be designed to best promote collaboration for teaching, learning, research, and outreach in the Higher Education environment?*

Some Examples

Some brief examples identified during the course of the VIUS project illustrate what some stakeholders wish to do but do not feel there are adequate tools available:

- Archeologists wish to share site data including images, graphics and other findings with colleagues at other institutions in a secure way
- Faculty and staff developing course materials held in personal collections wish to share with selected colleagues at their own and other institutions
- Faculty in disciplines such as architecture, landscape architecture, and agriculture wish to collaborate to build private departmental collections of images to be used for a variety of teaching, research, and outreach purposes
- Architectural historians are currently pooling their digital photography via the Society of Architectural Historians and seeking a sustainable way of continuing and expanding their efforts

- Faculty with scientific data sets wish to share with other researchers across geographically dispersed IT environments
- Students, faculty, scholars, researchers, and librarians wish to have a single search mechanism to locate learning resources held in a variety of distributed and centralized repositories

Project Description

There is a growing interest from commercial, research, and academic institutions to address these problems by developing scaleable models to share resources (images, course materials, data sets, etc.) and collaborate across disparate networks including traditional network infrastructures, emerging P2P networks and grid architectures. Furthermore, these communities wish to access large networks of resources from a single query to find resources to meet their particular needs. The proposed LionShare project hopes to address key issues of scalability, interoperability, and security by integrating, leveraging, and extending existing development efforts to achieve this purpose. The project team envisions an effort to address these issues by developing, implementing, testing, and assessing promising nascent work currently underway in several communities of practice including the LionShare development team at Penn State, the Internet2 Middleware Initiative Shibboleth project, the Open Knowledge Initiative (OKI) at MIT and the Industry Canada EduSource learning object repository (LOR) network efforts. Thus, the LionShare project team foresees this project having four major parts: 1) *the continued development of the LionShare P2P network infrastructure into a stable and scaleable software release*, 2) *the integration of middleware architecture for authorized access control developed by the Internet2 community as a means to federate P2P networks*, 3) *the development of middleware technology to connect and extend the LionShare P2P network to other central and distributed learning resource repositories using the eduSource architecture*, and 4) *the collection of requirements and evaluation of usability characteristics through planned assessment efforts*. Let us explain in more detail.

1.) *LionShare P2P Network Development*, will focus on continued development of the LionShare P2P architecture. It aims to leverage and build upon the P2P prototype development activities that were part of the VIUS project to create a stable, scaleable, easy-to-use software release for general distribution. It also leverages the considerable user analysis effort as part of the numerable stakeholder interviews, surveys, and focus groups during VIUS and integrates this into the functional requirements for software development. This portion of the development will add considerably new functionality for the end users of the LionShare Peer, administration features for PeerServer and added capability to the LionShare Networks. Penn State's Information Technology Services organization will lead this effort with the cooperation of the Internet2 P2P Working Group and the eduSource Canada organization.

2.) *Federating P2P Networks*, will address the need for a middleware infrastructure to allow sharing of resources ***between*** HEIs. Currently, the LionShare architecture is a ***single*** institution solution. Many instructors, scholars, researchers, and librarians wish to discover and share resources with colleagues at other institutions. This portion of the project will address this need by leveraging the authentication and authorization work of the Internet2 Middleware Initiative on the Shibboleth and the Open Knowledge Initiative (OKI) on Open Service Interface Definitions (OSID). This research involves extending the efforts of the Internet2 Shibboleth to support new models for authorization based on the trust architecture supported at the application level. This project will be led by a Penn State programming team, supported with consultation from the Internet2 Middleware and OKI community.

3.) *Connecting Learning Object Repositories*, aims to address the need to create bridges between multiple types of repositories, both central and distributed, for the retrieval of appropriate resources by users of P2P networks. It seeks to address the need to connect disparate networks to gain access to the collective resources of an international community of learning resources stored in public centralized collections (MERLOT[28], CAMEO, EdNA, VIUS, etc.) and distributed resources on developing P2P networks (LionShare, SPLASH, etc.). An eduSource Canada group will lead this effort with support from OKI and in collaboration with the Internet2 community.

4.) *Project Assessment*, will focus on a variety of assessment activities to verify and direct the software development process. Key to this effort is to craft applications that are easy-to-use and that meet the

requirements of the various stakeholders (individual and departmental collectors, collection administrators, and IT support staff, learning object repository communities, etc.). This assessment effort will be directed by Penn State and will be informed by the Internet2 P2P, and the Middleware Working Groups, and the eduSource Canada community.

A major part of the LionShare project is to build a sustainable P2P technology infrastructure through collaboration within a community interested in the technology agenda and the broader benefits that participation brings. All technologies will be developed under the Open Source GNU GPL license so technology developments can be highly leveraged in years to come for other purposes. We believe these development efforts provide significant opportunities for reuse in a variety of academic and non-academic areas and will contribute to the growing interest and development of secure P2P networks inside and outside academia. Cooperating groups committed to this project, thus far, include the Internet2 P2P and Middleware Initiative Working Groups, the eduSource Canada community, the Canadian Learning Object Repositories Networks (LORnet) group—now being formed, and the Open Knowledge Initiative. The Centre for Educational Technology Interoperability Standards (CETIS) in the UK has also expressed interest in participating.

These collaborative development efforts provide a significant opportunity to advance the state-of-the-art in the development advanced P2P networks, multi-institution access control, and multi-national learning object repository interoperability.

Section 2 – Background

VIUS and LionShare

The Visual User Image Study (VIUS), a 26-month Mellon funded project ending September 2003, is discovering how digital images are used for teaching, research, and outreach. Major components of the project include extensive interviews with stakeholders (students, faculty, and collections managers), surveys, focus groups, transaction logs, and protocol analysis to determine how individuals and groups are currently using images, where they find such images, and what they would like to do with digital images now and in the future.

Through this extensive contact with stakeholders, the Visual Image User Study (VIUS) team identified a number of criteria that point to the need for more flexible, user-controlled tools and services for discovery, organization, management, and sharing of personal and public images Many of the kinds of features that stakeholders identified suggested a much different model of collaboration was needed—one that provided much more control in the hands of the end user than available in large central repositories and that peer-to-peer (P2P) technologies offered many of these features. The VIUS team determined that peer-to-peer technologies offer unique opportunities to enable greater access and use of personal and publicly available image collections and that a prototype is warranted. The study also revealed the need for a more secure approach to ensure long-term viability and to protect the interest of personal collections. The study further determined that the vast majority of digital images reside in private collections and that access to and integration of public image collections is problematic.

The LionShare prototype evolved from these broad assessment activities, which were part of the VIUS, as a potential solution to these problems. It is this prototypical work that this proposal seeks to extend with further development activities. It also seeks to extend the definition of a learning resource beyond digital images to include a wide variety of digital resources including images, audio, video, presentations, simulations, research papers, Web resources, other learning resources and activities.

VIUS Assessment Overview

As summarized above, insights gained from the assessment components of VIUS have been a significant catalyst for the development of LionShare. Because we find the VIUS conclusions about the value of further development of LionShare (or similar sophisticated peer-to-peer approaches) so persuasive, it is important to be as clear as possible about the nature, scope, and thoroughness of that assessment.

The VIUS project was unlike many research and educational improvement projects – in which evaluation is treated as something of an afterthought – in that it was fundamentally based upon assessment. The VIUS project, from start to finish, incorporated frequent, varied, and carefully designed interaction with users: to explore needs, to involve faculty and students in the shaping of prototypes, and to assess how well products worked or didn't work from a customer's perspective. If the reader can think of a relevant digital-image university user (students, faculty, staff, librarians, instructional designers) or an appropriate assessment methodology (focus groups, surveys, interviews, transaction logs, think-aloud protocols), Penn State's project team probably incorporated it into VIUS.

The evidence underlying our belief that LionShare is a promising path for continued development draws from the following assessments, conducted as part of the VIUS project:

- Web and paper surveys of faculty and students in 68 Penn State departments of the arts, humanities, and environmental studies (with 1,473 total responses).
- Ten focus groups and 22 individual interviews with faculty, undergraduate students, graduate students, and staff from those same departments and from the Penn State University Libraries; those sessions included demonstrations and discussions of various prototype digital-image management systems (including peer-to-peer).

- End-of-semester paper surveys of students in two large classes that incorporated VIUS-provided software into course syllabi (with 462 total responses).
- Twenty think-aloud observational analyses of individuals using two image-management interfaces (one from the University of Washington, the other developed by the VIUS team) to complete an in-box exercise.
- Analyses of transaction logs for two image databases (the AMICO library of museum works of art, and the AP Multimedia Archive) licensed by the Penn State University Libraries; these data were based on user authentications tracked for over a year.

Some individuals participated in more than one of these assessments, and others not at all. Our best guess is that about 800 faculty members (out of 2,134 in the targeted disciplines) and about 1,100 students (out of about 21,150 in the targeted disciplines) participated actively —"actively" meaning that these counts exclude data extracted unobtrusively from electronic transaction logs.

VIUS Assessment Highlights

Three ideas justify the further development of LionShare. First, there is already strong interest in the organization, management, sharing, and discovery of digital image resources but few tools available. Second, there is a need for more effective tools, and that need is likely to grow. Third, the identified priorities of users appear more congruent with user-center approaches, such as sophisticated, authenticated peer-to-peer networks, than with more tightly administered and traditional repository management systems. These three notions are based on the following evidence from the VIUS assessment:

- *Students and faculty in the VIUS target populations are heavy image users* —75 percent of faculty and 55 percent of students studied use images in some form in their academic work. There is a wide range and variety in the types and images used, especially for faculty across disciplines, but on average faculty use about 1,100 images per semester and students use about 90.
- *Faculty, students, and academic departments maintain significant image collections* —Thirty-two percent of faculty said that they maintain personal image collections for academic use; the median size of faculty collections was 500 images. Students averaged about 50 images in their personal collections. The VIUS team paid special attention to so-called "intense users" — the 300 or so faculty members in the top quartile of faculty image users. Those individuals maintain personal collections of 2,000 to about 10,000 images. Image-intense departments at Penn State manage collections of as many as 350,000 images. For example, one faculty member in the intense user group in Landscape Architecture said he managed about 10,000 images, in large part for use in large-section undergraduate courses. Another faculty member in Anthropology likewise has been slowly but steadily converting his collection of 5,000 analog images, almost all slides, for use in teaching and research.
- *The trend is toward greater reliance on digital images* —At present, use is about evenly split between analog and digital images. But looking ahead, 75 percent of faculty and students say they plan to increase their use of digital images. Only 25 percent of faculty, and 33 percent of students, plan to increase their use of analog images.
- *Faculty and student use of formal, centralized, licensed databases is surprisingly light* — For example, the two databases tracked for the VIUS project contain in total about 765,000 images that should be of artistic, historical, and current interest to the 23,000 students and faculty targeted for this study (not to mention the rest of Penn State's 16,000 faculty and staff or its 83,000 students). They are readily available over the Web at no cost to anyone with a Penn State affiliation. However, on average, only about 100 people per month authenticated on either or both of these two databases. Even with availability of high-end centralized repositories with dependable image quality and populated, reliable metadata structures, these databases do not appear to be heavily utilized.
- *Image retrieval is more important than integrated software capabilities* — Reviews of existing software (including *Insight* from Luna Imaging Inc;[3] *MDID*[4] from James Madison University; and *CONTENTdm*[5] from DiMeMa, Inc.) and a peer-to-peer prototype clearly showed that users prize

image retrieval. If necessary, users are willing to sacrifice overall functionality of the application for a system that has a strong image search-and-retrieval capability. In fact, faculty in the "intense-user" category — a group that the VIUS team found to be especially perceptive —actually preferred the ability to concentrate on the discovery process rather than other aspects of image management.

- *Content is king* — If the findings of the VIUS assessment could be reduced to a single statement, it would be "Content is king." On several surveys (with responses from about 1,300 individuals in total) faculty and students consistently said that "access to a large number of images" would be their top priority from a list of 12 aspects of digital image management. Also, all indications are that users expect access to a wide variety of object resources, both inside and outside of Penn State.
- *Content needs are dynamic, not static* —From paper and Web-based surveys and, particularly, from focus groups and interviews, it was found that the interests of students and faculty are not static, but continually changing. Therefore, satisfying content needs will require an image delivery system that can respond quickly to changing needs and maximize access to the large and changing universe of available digital resources.
- *A successful system must address intellectual property concerns and metadata needs* —Again, from both surveys and focus groups, we have learned that closely following access to content as a priority are concerns about copyright, permissions, and related concerns, as well as the adequacy of metadata descriptions. Tools such as LionShare will need to meet such needs more seriously than have popular grass-roots peer-to-peer systems to date.
- *High image quality is not a deal-breaker*—Results of both surveys and focus groups consistently have shown that faculty and students do not expect to rely on virtual image retrieval systems for publication-quality digital images. That is not the primary need for the bulk of teaching and learning applications. For example, one survey of 700 faculty respondents brings to light that over four times as many individuals were concerned about whether "the content of the images collected suits my areas of interest" as about whether "the quality of the digital images will be inadequate."

In conclusion, user needs carefully and systematically identified in the VIUS project strongly supported the strengths of technologies designed to help members of extended virtual communities pool and coordinate large sets of distributed resources. Equipped with these findings, the VIUS team envisioned a prototype entitled LionShare, based on emerging research and development related to P2P networks. The following section provides a description of the philosophies that emerged to lead the development effort.

LionShare Philosophy

LionShare has developed several philosophical underpinnings that emerged out of our conversations with stakeholders. These philosophies are reflected in the design of LionShare:

- *Personal Information Management* —With the proliferation of all types of digital objects and the growing storage capacity on the personal desktop, it was recognized that new easy-to-use personal digital management tools were needed on the desktop. Tools that are primarily used for organizing personal collections but could be used to connect to public image repositories when necessary.

- *Simple, Intuitive User Interface* — Because of the vast amount of information that can be stored, new strategies for discovery and use are required that greatly simplify the process of discovery, organization, management, and sharing. These tools must leverage the increasing computing power on the desktop and advances made in standards for federated search and retrieval in the digital library community.

- *User Defined Sharing* —The holders of image resources need to be responsible for the scope to which they share personal collections. Tools need to be developed to make it easy for the owner to setup file-sharing privileges at their own discretion. This would allow owners to manage personal collections and have the choice to contribute or connect those resources to local, regional, or public collections.

- *Authenticated Access to Network* —The need for accountability for distribution of digital resources was noteworthy and should be recognized. Thus, the use of file sharing networks needs to be connected with an individual sharing a digital resource. By incorporating domain-specific authentication mechanisms, like Kerberos, files can be directly attributed to a specific file sharer. It was recognized that authentication also provides an important benefit in that it provides the ability to access information about the sharer from central services, like LDAP[7]. This functionality allows for the use of such information to automate the process of metadata creation. This information could then be incorporated into the resource description.

- *Standard Metadata Structure* —Standardized descriptive information needs to be associated with each resource being shared to increase the likelihood for discovery and reuse. Likewise, the time and effort required to create metadata descriptions need to be greatly reduced by using self-organizing strategies. If Content is King, then Metadata is Queen. In other words, without a standardized metadata structure discovery becomes much more difficult. The standards for learning objects have evolved considerably over the past five years and these efforts have been incorporated into LionShare.

- *Leveraging Open Source*[8] —The advantages of Open Source efforts were recognized as an economical way to leverage the software development activities of others in an effort to realize new potential. Open Source communities distribute the development effort over a large group community interested in common goals. Using Open Source could greatly enhance the time to market a new software initiative by starting with an already available code base. Open Source development technologies such as CVS[9] make it easier to expand the LionShare development team and allow for co-development beyond the projects lifetime.

Prototype Accomplishments

With these overarching philosophies in mind, a development plan was created for an initial alpha release (v0.2) having a very limited distribution and scope, which has served as a proof-of-concept. This development was enthusiastically received in early stakeholder demonstrations and presentations. The success of this alpha research encouraged the LionShare team to develop a more ambitious development plan for the initial alpha release adding more requested functionality and requiring further user testing. During alpha development, user testing and feedback have been extremely important. Feedback has been collected and used to improve the release and solidify the feature set.

Furthermore, a beta (v0.5) release has been outlined with additional capability. This release is currently under development and will incorporate user feedback from the July alpha testing. The beta will provide a number of additional new features that were not in the alpha release but can be completed within the scope of the original VIUS grant period (September 2003). The beta release scheduled for September would demonstrate the basic functionality without the final look-and-feel required for a final production release.

Additionally, a development plan for future work beyond the scope of the grant has also been created to document future effort needed to move LionShare to a production release and incorporate some of the requirements being articulated by I2, eduSource[10] and OKI communities. Further refinements would be necessary for a finished production release (v1.0) of LionShare and substantial research, testing of the scalability of the basic architecture. The following section describes the Basic Architecture —a centralized+decentralized topology[11].

Basic LionShare Architecture Outlined

The basic architecture for the LionShare network includes three elements: the Peer, PeerServer and supporting networking. These three elements are briefly described below and represented in *Figure 2.1. The LionShare Topology*.

- *LionShare Peer*— In the P2P world, the peer functions as both a client and a server, but in the case of LionShare (a hybrid topology) a federated P2P with a decentralized+centralized topology is used. This structure allows users to post files and/or metadata to a centralized, persistent server, instead of having to run the application locally, one-to-one, on their system at all times in order to share. Users can still share locally, but the federated approach adds convenience and persistence to the architecture. The LionShare peer is a heavily modified version of the Limewire Open Source project[44]. Limewire is a multi-platform P2P application written entirely in Java. The source code for Limewire is distributed under the GNU GPL[12] license and may be freely distributed as long as it remains Open Source. Federation is not the only feature that differentiates LionShare from Limewire: authentication, multimedia support, automated metadata creation, and so forth.

- *LionShare PeerServer*—The LionShare PeerServer is at the center of our enterprise approach to P2P, functioning as a file server for LionShare users. The PeerServer looks like any other LionShare peer except when a user initiates an upload request. Users have the ability to upload and manage files and metadata located on the PeerServer. When queries are initiated on the LionShare network, files stored on the PeerServer are listed in the same manner as files shared locally running on the LionShare Peer client. The LionShare Peer provides a number of other functions for managing, organizing, viewing, and sharing images.

- *LionShare Network* — The LionShare network remains basically unchanged from the Gnutella protocol specification[13]. Additional functionality for authorization was added, but to the application itself, the PeerServer looks like any other peer except when a user initiates an upload request. When a peer connects to the LionShare network, it attempts to connect to several different IP addresses of known Gnutella hosts. A host can be any peer running on the network but usually only high-bandwidth peers, known as Ultrapeers are used as hosts via a connection preferencing algorithm.

  LionShare uses two methods to find hosts for making a connection: the first method is by the use of a Gnutella *host cache*. Each peer keeps a dynamic list of possible hosts by pulling Gnutella IP's from regular connection traffic. The second method for acquiring host addresses is by the use of a Gnutella *Web cache*. A Gnutella Web cache is a dynamic PHP[14] script that can be located on any Web server that supports PHP. The cache is a database of 10 Gnutella Ultrapeers that are currently active. Gnutella clients use the addresses of those Ultrapeers to connect to the network. The LionShare network takes advantage of both of these connections to privatize our network and customizing how the host cache and the Web cache are used within LionShare.

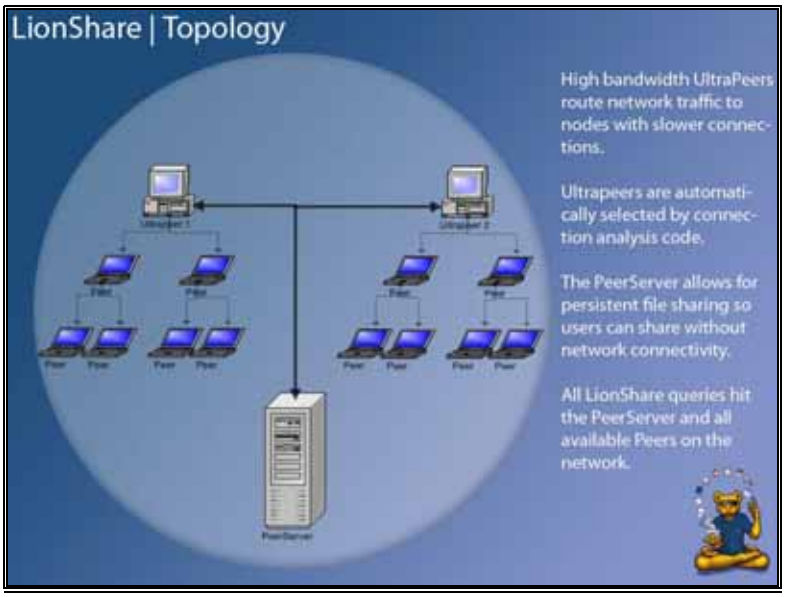


Figure 2.1 LionShare Topology

LionShare Centralized+Decentralized Topology

As mentioned earlier in this document, LionShare uses the Gnutella protocol as the basis for its networking. The original Gnutella protocol specification was completely decentralized, meaning all peers were treated equally on the network. There are many issues with a completely decentralized topology, but the primary issue turned out to be scalability. Modem connections were treated the same as broadband and T3 network connections. The original Gnutella network was easily bottlenecked by a few modem connections. Network queries would die instantly when attempting to pass through a peer on an overwhelmed low-speed connection. To solve this issue, Gnutella developers came up with a hierarchical topology. The Gnutella network would still be decentralized, but lower bandwidth connections would pass through high-speed gateways known as Ultrapeers.

Early on in the LionShare project, we noted that one of the downsides of P2P file sharing was the requirement of being connected to the network and running a background application in order to share. A partial listing of groups who are incompatible with such a requirement includes: users on the road, users who have low speed connectivity, and users with slower computers. Another issue that came to our attention was motivation to share with educational related content. Potential users may be less inclined to keep educational sharing applications open compared to more “general purpose” Internet file sharing applications. In order to address these concerns, it was decided early on to add an element of centralization to the LionShare topology.

The topology of the LionShare network is best classified as a centralized+decentralized or “hybrid” topology. The same hierarchical decentralized P2P topology exists, but we added the element of a PeerServer. The LionShare PeerServer acts just like a traditional file server; users can upload files and/or metadata onto the PeerServer. Those files uploaded to the PeerServer are shared, even if the user decides to close the application and disconnect from the network. Another advantage to this topology is that the PeerServer performs services for the network. For example, a lightweight authorization engine might be added to the PeerServer, which would allow users to upload files, but they can decide to deny access to all. The authorization engine can effectively turn a PeerServer into personal remote backup tool.

LionShare Benefits

This basic centralized+decentralized topology provides us with a variety of benefits for creating private networks for collaboration. The following section outlines a few of the benefits of the LionShare architecture:

- *Allows persistence to the P2P architecture* — One of the disadvantages often recognized with P2P architectures is the lack of persistence of metadata and resources. When a resource holder is not on the network neither are their resources. In the LionShare architecture, the PeerServer can act as a local aggregator of metadata, object resources or both. While it is still the end-user that controls what is being shared and how widely, the PeerServer, when given the authorization can hold and provide a persistent location for the resource to be shared.

- *Ability to create a private network for group collaboration* — Because the PeerServer can be configured in a variety of manners, it is possible to build small private networks with access limited to a small group. Access control can be limited by either a specified user group or by IP address.

- *Opportunity for federating multiple smaller networks as needed* — Should there be a need to create larger networks, smaller private networks can be connected together to build increasingly larger networks. Additionally, it is possible for a P2P network to be connected to a central client/server resource through the use of shared metadata structures.

- *Ability to use local authentication methods to gain access to the LionShare network* —One of the unique aspects of the LionShare network is the ability to authenticate a specific user of the network against an authentication service. This authentication process makes it possible to control access to the resources of the LionShare network. It also provides a mechanism to track how resources are used on the network.

- *Ability to build specialized application around this architecture*—The core functionality of LionShare provides the potential to develop a variety of specialized applications on top of this functionality.

Section 3 – Strategic Goals

LionShare Research Questions

The primary research question to be answered in this project is:

- How can P2P technologies be designed to best promote collaboration for teaching, learning, and research in the Higher Education environment?

To answer these questions a number of more specific questions need to be addressed that are user-centered, community-centered, or architecture-centered.

User-centered questions

- What is the right user interface for end-users?
- What is the right user interface for server administrators?
- What additional features are required?

Community-centered questions

- How are communities of practice supported?
- Can P2P technologies support large federated inter-realm networks?

- How is LionShare marketed?
- Who is the audience for LionShare?
- How valuable is the ability to bridge between independent communities?

Architecture-centered questions

- What is the right architecture to provide a robust, scalable network?
- Can P2P technologies be made secure, scalable, and user friendly?
- How can the architecture better support both Grid and P2P network users?
- What interoperability standards should be supported?
- Can access rights control systems (such as Shibboleth) be used for crossing between heterogeneous networks connected via gateways?
- Can two protocols/APIs addressing a same problem at different levels of abstraction be smoothly integrated?
- Can heterogeneous networks be smoothly connected? How can such a system be kept open and expandable?

Section 4 – Peer-to-Peer in Academia

Peer-to-Peer for Academic Networks

When people today hear the acronym P2P, most commonly associate the technology with the trading of music, movies, and other copyrighted material. LionShare is drastically different than the popular file sharing tools used on the Internet today, and it is important to note that file sharing is just one of many P2P technologies. Instant messaging, Web services, and distributed computing are just some of the technologies that fall under the P2P umbrella. Collaboration and organization are the two concepts central to the design of LionShare. File Sharing is just one technology that falls under the category of collaboration. The vision for LionShare is one that centers on giving users the ability to share, search, download, and organize digital media while providing collaborative tools for individuals, departments, and organizations to communicate, collaborate, and share. When the idea of LionShare first arose, many people were intrigued, but quite a few had one simple question. Why bring P2P to academia?

The Digital Media Age

Sometimes it is hard to remember how far technology has come in recent times. As little as 15 years ago, learning environments were drastically different than what we see today. A high tech classroom consisted of a television and an overhead projector. Assignments were handwritten or in some cases machine typed. Computers were still very much a novelty for the average student, and most could not afford the $3,000 price tag. The quickest way to get a photograph was with a Polaroid camera. The Internet was a text only world for a few lucky scientists, engineers, and researchers. A fax machine was state-of-the-art communications technology, while students used large volumes of books called “encyclopedias” to do basic research.

Today, there are some college freshmen that have never even touched an encyclopedia. High-tech classrooms feature wireless Internet, high-resolution digital projection, and course management Web sites. Students send their assignments and collaborate on group projects via the Internet. Class presentations are now referred to as “Powerpoints,” while most research takes place over the Internet. Digital imagery has a strong foothold in higher education. Google image search and the ever-decreasing price of digital cameras are two of the many reasons for the large growth numbers of digitized imagery. Digital video and audio are now commonplace in the classroom.

Many of the digital media technologies and file formats we use today have existed for over five to 10 years, but the difference today is the ever decreasing barriers to entry. Digital media is now affordable thanks to economies of scale and continued innovations. One barrier to entry used to be storage. The price of IDE hard disk storage has slipped to well below a dollar per gigabyte. Affordable two mega-pixel digital cameras are available for $100 and digital video (DV) cameras hovering around the $300 mark have made digital media creation accessible to the average student. Personal digital media collections are now a mainstream phenomenon. Digital photo, audio, and video editing are now common thanks to Moore's law and the fact that all the major consumer operating systems now come with basic digital editing tools. Lightweight editing software is typically bundled with consumer level devices.

Scenarios: Current Practices vs. LionShare

Digital media has a strong foothold in nearly all aspects of higher education. Students, faculty, and staff are regular users of digital media; in fact many are becoming overwhelmed by it. Acquisition, organization, and distribution are three of the main problems associated with living in the digital media age. LionShare is a tool that addresses all three problems, but it is important to understand some of the issues associated with digital media in higher education. One of the key reasons why LionShare was envisioned early on was to eliminate redundancy in the act of acquiring digital media. The following examples are typical situation experienced by faculty and staff, and how LionShare would address those scenarios.

Faculty Example #1

Two professors in the same department (botany) are assembling their course materials for the upcoming year for related classes. Both are looking for similar digital images for use in various Powerpoint slides during the upcoming semester. Instead of finding out that they are both looking for similar items, both professors end up spending an inordinate amount of time doing the same amount of work. There are also three other faculty members who taught the same class previously and spent the same amount of time for looking images.

Current Practice
This is a yearly ritual that is performed by thousands of new faculty across the globe, the acquisition of digital media. Someone else also found most of what they are looking for, either in the same department or in a related field. Most faculty would be happy to share their media collections if they knew there was a need. The faculty in need of media would be happy to ask for the media if they knew of some other faculty member on campus who already had it. Personal collections are another issue, as examined in the following example.

LionShare
With LionShare, a faculty member could easily share any digital media, making it accessible to people with similar interests. It would also be very easy to find other people on the network with common interests. Sharing files would be extremely easy and if one did not like the idea of leaving an application running on their computer, they could simply upload the files to a PeerServer. The main benefit of LionShare in this case is reducing the amount of man-hours required to assemble media for lesson plans and encourage discussion amongst people with similar interests.

Faculty Example #2

A landscape architecture professor happens to be a photography enthusiast. Over the past few years she has acquired a large collection of digital imagery of famous landscape designs professionally shot during her travels. She uses much of imagery for classroom examples but only a few other department members are aware of her collection.

Current Practice
Typically, a personal collection like this could be used by a great deal of faculty, and the people who amassed the collection are only happy to share. The problem is typically a matchmaking issue. In some cases there might by a department file server, where faculty can dump media in to a directory. Unfortunately, a great many faculty do not have the technical savvy to post or acquire information shared on a file server. The other issue is organization of imagery; file systems typically do not contain any meta data and the servers are restricted to only department members. People looking through a file server have to rely on filenames as a descriptor for the media. Another distribution method for personal collections might be a Web site, but generally the technical knowledge required is too great for the average user and the process can be time consuming even for a technically savvy faculty member. The next two examples describe how students are using digital media.

LionShare
LionShare would allow the professor to easily share her photography collection with other department members and anyone else with access to LionShare, depending on to whom she granted access. Using the built in metadata entry tools, the professor could fill in the image descriptions individually or use the bulk metadata tool to fill in common fields across multiple files.

Student Example #1

An International Marketing class has a new assignment that splits the students up into groups of three to do a presentation that requires Powerpoint. The group of students exchange e-mail addresses and try and decide on a time to meet to pick a topic and distribute the workload. All three students have very busy schedules and quickly decide that they will just manage the project through e-mail. After one week of e-mails they decide on a topic and a workflow. One person is assigned editorial duties, putting together each

portion. On the day before the project is due, the person assigned editorial duties receives two e-mails. One e-mail has a notice stating that the 5 MB e-mail exceeded the attachment limit. The second message has an attachment but it appears to be corrupted.

Current Practice
Students typically use e-mail as the preferred method for file distribution. There are many reasons why e-mail is a terrible file distribution method. E-mail attachments have a tendency to become corrupted. Even if the file goes through a maze of SMTP servers unharmed, there is no guarantee that the recipient will be able to open it, especially on a different platform. Typically, a student group finds out about these problems after they have the time to do anything about it.

LionShare
For group projects, students could log into LionShare and create their own private chat room. They could add the user ids of their friends to alert them when they appear online. If one member is not going be online to share a file, he can upload the file to the PeerServer and restrict access to the editor's user id only.

Student Example #2

A student is taking a semester abroad in Barcelona. She can't wait to meet new people and experience a new culture. She is also excited because her parents just bought her a new digital camera for her birthday. While on the trip, she decides to send some photos of her new friends to some classmates back home. She includes the JPEG images with an e-mail but the message is not sent, instead she receives a cryptic error message.

Current Practice
Typically, digital cameras use high resolution and low compression levels by default. Digital camera novices have no idea how to reduce image size and increase compression levels to make images suitable for Internet distribution. Most e-mail systems have tight limits on attachment sizes in order to save costs and to discourage users from using e-mail for file transfer. Because of the increasing popularity of digital media among students, university systems and network access are now being used for personal file distribution. In the next example we will look at an access control scenario.

LionShare
Anywhere a user has access to the Internet, they can connect to the LionShare network. Sharing personal collections with others is a perfect use of P2P technology. With traditional P2P networks, the student would be forced to leave her computer on in order for her friends to download but this method would be inefficient and time consuming when abroad on a metered connection to the Internet. LionShare would let the student upload the files once to a PeerServer and disconnect from the network. In this scenario the PeerServer saves time and money.

Authorization Example #1

A staff member is working on a proposal for a grant to the Mellon foundation. The proposal includes a large amount of diagrams, boosting the file size to over 10 megabytes. The staff member would like to share the proposal with a handful of colleagues but would like to restrict access to those individuals only.

Current Practice
This is a common scenario where there are several possible channels for distribution, but none are optimal. E-mail is ruled out because of the file size. The department file server is only accessible inside a firewall and even then the staff member could not specifically access by user name. The common method used for distribution in this case is uploading the file into a folder on a Web server. Because the folder/file is not linked directly from a Web page, users mistakenly think that no uninvited guests will discover the directory. Unfortunately, the address to this directory is usually sent via plain text e-mail and there are several different Web server scanning techniques which can uncover non-linked directories.

LionShare

LionShare's authorization system would allow users to grant or restrict access to specific groups, individuals, or universities. For confidential files, users could manually enter user names to grant access. Using LDAP information, files could be restricted to members of a certain group or department.

Federation Example #1

Faculty member 1 from University A decides to collaborate with faculty member 2 from University B on a new research project. Faculty member 1 already has a large amount of documents and media published on a Web site. Unfortunately, faculty member 2 cannot access the Web site because it authenticates off university A's Kerberos server.

Current Practice

Faculty member 1 has a few options at his disposal. He could contact the university's computing services and set up a restricted user name for his colleague at university B, but that could take weeks to go through. He could save all the documents from the Web site and e-mail them, but that would be time consuming and some of the files are just too big. Typically, what happens in such a scenario is that faculty member 1 sends a plain text e-mail with his account information and password, making his information available to anyone who happens to be sniffing tcp/ip traffic.

LionShare

LionShare's authorization system would give users the ability to share files with other participating universities and restrict access with same options listed in the authorization example in the previous section.

Federation Example #2

A oceanography faculty member is preparing a lesson plan for an upcoming class, she is trying to find a set of animated video clips that illustrate the effects of sea floor spreading. She knows they exist because she ran across them on a educational resource website a few months before.

Current Practice

There is simply not one-stop place to find educational resources. Today, the faculty member would probably try using Google and if her Google search was unsuccessful, she would probably spend an inordinate amount of time checking multiple repositories for the specific clips she was looking to use.

LionShare

It is very possible that if the faculty member in question had a tool like LionShare at her fingertips from the beginning she would have added any interesting academic content to LionShare the moment she came across it. In the example description above, it is assumed that she was not aware of LionShare when she first viewed the animations a few months before she needed them. If she did not save the clips, she could query the LionShare network to see if any of her colleagues managed to save animations illustrating sea floor spreading. Because LionShare is connected to other networks, for example fixed repositories, she could search multiple academic P2P networks and repositories. With LionShare she would only need to do one query, instead of using the current methods of checking multiple websites or searching through loads of unrelated content using a general search engine like Google.

Section 5 – Project Description

Introduction

There is a growing interest from commercial, research and academic institutions in the development of the scaleable models to share resources and collaborate across disparate networks including traditional network infrastructures, emerging P2P networks, and grid architectures. The LionShare project hopes to address key issues of scalability, interoperability, and security by integrating, leveraging, and extending existing development efforts to achieve this purpose. The project team envisions an effort to address these issues by developing, implementing, testing, and assessing promising nascent work currently underway in several communities of practice including the LionShare development team at Penn State, the Internet2 Middleware Initiative[15] Shibboleth project[16], the Open Knowledge Initiative (OKI) at MIT and the Industry Canada EduSource learning object repository (LOR) networks efforts. Thus, the *LionShare: Connecting and Extending P2P Networks* project team foresees this project having four major components: 1) *the continued development of the LionShare P2P network infrastructure into a stable and scaleable software release*, 2) *the integration of middleware architecture for authorized access control developed by the Internet2 community as a means to federate P2P networks*, 3) *the development of middleware technology to connect and extend the LionShare P2P network to other central and distributed learning resource repositories using the eduSource architecture*, and 4) *the collection of requirements and evaluation of usability characteristics through planned assessment efforts*.

Shibboleth

The LionShare team would like to see their future work aimed at federated, authorized P2P network using the MACE Shibboleth[17] and OKI architecture. In layman terms, this would enable an authenticated P2P network at one institution to search and access resources at another institution using a trust model between institutions. Of particular note, is the growing interest within the Internet2 community in using the LionShare P2P network as a test-bed for developing, testing, and deploying a security infrastructure for authorized access control for federated P2P networks. The federated P2P efforts proposed in this project explore new territory for security infrastructure research, in that it seeks to address how access control can be enabled without using Web Service architecture. This federation process would provide the ability to connect multiple P2P networks into one loosely coupled authorized network despite the use of different local authentication technologies. The ongoing development of open inter-institutional authentication technologies will eventually lead to a revolution in resource sharing amongst institutions. In the past, universities and companies looking to share information over the Internet could not do so because of the need for a common authentication architecture that would allow multiple yet separate institutional end user identities. Shibboleth, a project from Internet2's MACE group (Middleware Architecture Committee for Education), is an exciting new architecture that could provide federated P2P while maintaining security, user identity, and information.

OKI

Of additional interest, is the awareness within learning technology community of the importance of the integration of OKI Open Service Interface Definitions (OSIDs) into the development of the software architecture for LionShare. Higher education leaders recognize that learning environments are a core component of their information technology infrastructure and that any supporting technology, like LionShare, would benefit from leveraging the OKI OSIDs. The use of OKI technologies ensures the flexibility to adapt to changes in the middleware technologies without breaking the supporting technologies. The technologies must be robust and must scale up to support an ever-increasing demand. The project team anticipates that the OKI OSIDs for authentication, authorization, filing, and digital repositories will be key components of this project. During the course of the project the use of other OSIDs may find their way into the project.

EduSource Canada

Additionally, the Canadian EduSource community, part of CANAIRE[18], has expressed interest in working collaboratively on connecting and extending LionShare networks to include learning object repository networks in Canada. The Canadians SPLASH[19] P2P technology is part of the EduSource initiative that is

developing a network of learning object repositories that are both distributed and centralized. Of interest is how P2P networks using different protocols can be connected together to share resources. Additionally, the Canadian goal is to connect several types of repositories together using the IMS[2] Digital Repositories Interoperability (DRI) specification[20]. Middleware developed by I2, IMS and OKI has the potential to bring this vision to fruition. Furthermore, additional efforts are needed to develop common middleware architecture for the project that incorporates the requirements of all of these communities. LionShare partner information follows.

The following section describes the details of what each work group will undertake and the deliverables associated with each group's activities.

Part 1
LionShare P2P Network Development

Application Development

While great progress was made over the past 10 months, LionShare is still a prototype, and as such, requires considerable development activity to bring it to a production level release. In the following section, the features necessary to bring the LionShare to the level of a public release will be discussed. The following will be organized around the three basic architectural components of LionShare: Peer, PeerServer, and Network. Descriptions of each component will include the pre-funding status and the features planned for a version 1.0 release.

LionShare Peer

The LionShare Peer is the application installed on the desktop system to provide file sharing and personal information management tools. The Peer extends the interface provided by the Limewire code base. The application is written entirely in Java and is compatible with all major platforms or any system that will run a JRE[21] (Java Runtime Environment) 1.4 or greater. The LionShare Peer utilizes Limewire's intuitive tab-based interface providing ample room for additional functionality without overwhelming the user with complexity.

The alpha version of the LionShare Peer features automated metadata creation and file sharing tools. The interface includes three main tabs: search, library, and published. The *search tab* is where users can query the LionShare network, searching by media type, or querying the all media types shared on the network. Each query has its own tab, so users can initiate multiple searches simultaneously with efficiency. The *library tab* provides the media management capabilities of LionShare. From the library tab users can add resources to share, upload resources to a PeerServer, describe shared resources, or play media using internal and external media players. The *published tab* displays what files the user has uploaded on a PeerServer and allows users to remove files from a PeerServer. Additionally, there are two advanced tabs that are not shown using the default interface: monitor and connections. The additional tabs provide network information for curious or advanced users.

The LionShare Peer is the primary component that enables personal information management of digital resources. As such, the LionShare Peer needs to provide robust capability and tools for this purpose. It must create value on the desktop for the *LionShare* user. Additionally, it must have powerful and scalable tools for collaboration and sharing. Version 1.0 features will focus on expanding the user experience, providing productivity tools and networking functionality to enable this vision. The following functionality is envisioned for the *LionShare* desktop environment.

- *Setup and Installation*—A multi-platform installation tool will be chosen once to distribute the application for wider testing. There are a variety of open source and free applications that will work on all major platforms (Windows, Mac OS, Linux). There will be the usual default, minimum, and advanced installation options. The advanced installation option will give users the ability to add or remove features to be installed. The development of additional features for the LionShare peer will remain modular so LionShare can maintain a small footprint with the minimum installation option.

- *User Interface* — A variety of user interface features have been identified to improve the usability of the application. Customized skins will give LionShare a unique feel while maintaining the simple UI.

- *Metadata* — Metadata entry, automation, and import/export are the core components of the LionShare metadata strategy. The LionShare Peer will harvest as much information as possible from file systems and file headers. User profiles are the central part to our metadata strategy. User profile information can be used to automate metadata entry. Profile information can be manually entered but another option will be to connect user profiles to an LDAP (Lightweight Directory Access Protocol) service. The automation of metadata is crucial. Users have shown time and time again that manual metadata entry is not a viable option except for a small minority, especially when there are many possible fields. An import/export function based on standards will provide another source for entry and the transfer of information to other applications and learning management systems.

- *Multimedia* — LionShare already has support for basic image viewing (gif, jpeg) and support for mp3 playback. The Peer integrates with external media players, allowing for playback of multimedia using players already associated with the media type based on the assignment by the user's operating system. Additional internal viewer support will be added for other image and media types if the need exists. Support for other image formats and PDF files for the internal viewer is scheduled for the beta release in September 2003. Another feature expected in the September release is the ability to export a set of images to an HTML slideshow.

- *Remote backup* — When the authorization is finally implemented for the PeerServer, the ability to push files to a PeerServer for backup purposes will exist. Users can publish files in the library tab and modify the permissions to deny access to all other users. This could be yet another use for the persistence feature of the LionShare model.

- *Instant Messaging*[22] — Users have the ability to chat with others connected to the LionShare network but there could be benefits to supplying a more robust form of social negotiation. Instant messaging could be as easy as supplying an instant messenger address in the user profile or integrating a full fledged IM client in to the application if needed.

- *Chat Rooms* — A chat room feature could be a great way for LionShare users from common academic backgrounds to meet and exchange files. Integrated chat rooms in P2P applications are not new. The first versions of Napster included chat functionality similar to IRC (Internet Relay Chat). LionShare users could meet, exchange ideas, and browse other users' shared files from the chat screen. LionShare's centralized+decentralized topology would allow the PeerServer to function as a chat room server.

LionShare PeerServer

The PeerServer is at the center of our enterprise approach to P2P, functioning as a file server for LionShare users storing metadata and objects that are being shared in a persistent way. Users have the ability to upload and manage files located on the PeerServer. When queries are initiated on the LionShare network, files stored on the PeerServer are listed in the same manner as files shared locally running on the client.

Future PeerServer development efforts include extending the usability of ordinary file serving services to a robust administration tool for LionShare P2P networks. Systems administrators would have the ability to view statistics and change settings from a Web-based interface. Files shared on the PeerServer could also be downloaded via the Web and the PeerServer could also function as a remote backup tool, by giving users the ability to post files to the PeerServer in a private folder.

- *Architecture* — To ready the PeerServer for a production release, significant changes will need to be made to the design of the file transfer and storage components. Currently, the PeerServer relies on the same methods used when storing files locally using the Peer application. The use of an Open

Source relational database package will improve performance and scalability. The file transfer piece will also need to be optimized for scalability purposes.

- *Administration* — Currently there is no administration interface for the PeerServer. Development of a Web-based administration package will be needed to make the LionShare network a more viable production service. The overhaul of the PeerServer storage architecture will assist the development of an administration interface. The logical solution is to provide a Web-based administration interface where administrators could view activity, statistics, logs, and change settings from a Web browser.

- *Authorization* — A lightweight authorization engine will need to be in place in order for LionShare to be a viable production service.

- *Audit system* — In order to maintain LionShare's goal as an educational tool, an audit system will need to be place for administration and research purposes. An XML based log format would be the logical choice. Administrators and authorized researchers could view log information on the Web-based administration interface. Reports of statistics and suspicious activity could be e-mailed to them periodically.

LionShare Network

The LionShare network remains basically unchanged from the Gnutella protocol specification. Additional functionality for enterprise level uses were added, but to the application itself, the PeerServer looks like any other peer except when a user initiates an upload request.

When a peer connects to the LionShare network, authentication via Kerberos or other authentication protocols is required before the network connection routine takes place. Once authentication has occurred correctly, the Peer connects to several different IP addresses of known Gnutella hosts. A host can be any peer running on the network but usually only high-bandwidth peers known as Ultrapeers are used as hosts via a connection preferencing algorithm.

LionShare uses two methods to find hosts to connect to; the first method is by the use of a Gnutella *host cache*. Each peer keeps a dynamic list of possible hosts by pulling Gnutella IP's from regular connection traffic. The second method for acquiring host addresses is by the use of a Gnutella *Web cache*. A Gnutella Web cache is a dynamic PHP script that can be located on any Web server that supports PHP. The cache is a database of 10 Gnutella Ultrapeers that are currently active. Gnutella clients use the addresses of those Ultrapeers to connect to the network.

- *Authentication* — LionShare authenticates users via the Kerberos network authentication protocol. This is a satisfactory method for Penn State University, but other methods must be supported for the LionShare code to become usable on other institutional networks that utilize other authentication methods such as LDAP or PKI[23]. An authentication method for an inter-institutional version of the LionShare network is another direction for the LionShare project. Support for the Shibboleth middleware Web authentication project is outlined in the following section of this document on *Federated P2P Network.*

- *Multiple Protocols* — Currently, only a slightly modified version of the Gnutella protocol is put to use on the LionShare network. The Gnutella protocol has many advantages, but some disadvantages include a large amount of network traffic for queries and other protocol traffic, plus some scalability issues that need to be addressed. For an internal P2P network, such as a Penn State only P2P network, the scalability and network traffic issues are insignificant. However, an inter-institutional model does create quite a few challenges that must be considered. Traffic between institutions over the Internet provides a new set of requirements. Many optimizations to the Gnutella protocol have reduced the amount of bandwidth required for protocol messages, but it could still prove to be an issue for an inter-institutional network. Support for other protocols may be necessary as the requirements for LionShare change.

Project Organization: LionShare Application Development

The project leader for the LionShare architecture effort will be lead by Michael J. Halm, Senior Strategist, Teaching and Learning with Technology (TLT) group, part of Information Technology Services (ITS) at Penn State. Mike has played a key role in developing the LionShare architecture at Penn State. He will have technical support provided a Lead LionShare programmer, as well as additional programming support for software development.

In addition, liaisons will be assigned from the Internet2 P2P Working Group. This group will provide consulting and technical support, participate in co-development opportunities and provide a test-bet environment for the pilot project. Furthermore, this group will play a significant role in providing requirements for the P2P environment and testing of the LionShare P2P network. A call for participation will be issued to the Internet2 InCommon federation. The group will have a regularly scheduled conference call, minutes and a Web site supported by the Internet2 P2P Working Group.

Internet2 P2P Working Group

The Internet2 P2P Working Group (http://p2p.internet2.edu/index.html) consists of more than 400 members from research institutions, government, corporate entities, and others who share an interest in the P2P environment, its effects on the higher education community and its possibilities. The working group's mission is expressed in the following:

- To provide a forum for reporting on recent occurrences and future trends within the peer-to-peer and distributed computing space. The forum may occur at the regular member meetings, at Joint Techs meetings, or at specific workshops designated for the topic.
- To be a clearinghouse for collaborative opportunities within the higher education community and between that community and corporate entities as new peer-to-peer and distributed computing applications and tools are investigated.
- To provide best practices documents for both resource management as well as innovative uses of peer-to-peer technologies.
- To provide a central repository for resources and documents related to all aspects of peer-to-peer computing.

In previous and on going efforts the working group include: 1) evaluated a P2P collaborative environment tool, 2) developing a bandwidth management best practices document with the Campus Bandwidth Management (CBM) working group, 3) developed a sub-group who is exploring P2P security issues, and 4) provides an information point of reference on P2P concerns through its Web site.

Other P2P Technologies in Higher Education

The status of other P2P technologies within the Higher Education sector is somewhat cloudy. David Futey of Stanford University and Co-Chair of the Internet2 P2P Working Group reports:

> *At this time there appears to be no competing efforts within the US higher education environment, though projects at a local institution level may be in progress unbeknownst to the working group. Interest in participating in the Lionshare project was expressed by its member at the Spring '03 member meeting when the possibility of the project was announced. The working group sees significant benefit to the Lionshare project. The P2P environment has been sharply criticized for how the technology has been used which in turn has, to a degree, criticized the technology. The Lionshare project offers the higher education/Internet2 community an opportunity to develop an application that can reflect a positive and productive file-sharing framework within a P2P architecture. The P2P working group offers the project administrative support, feature testing, potentially programming assistance, and eventually a conduit for adoption in the community.*

While the P2P space in US Higher Education is still developing, there are several P2P efforts within the international community. The most notable are the SPLASH effort in Canada sponsored by CANARIE,

Canada's advanced Internet development organization and part of eduSource, our partner in this proposal, and the Edutella (http://edutella.jxta.org/) effort based in Germany. A brief description of each follows.

SPLASH

In 2000, the Portals for Online Objects for Learning (POOL) Project was launched under the CANARIE eLearning Program. CANARIE is a Canadian agency of government, industry, and academia to promote broadband infrastructure and application development. Dr. Griff Richards, then Research Integration Officer for the TeleLearning Network of Centres of Excellence, led Phase II of POOL. While Phase I of POOL focused on centralized portal solutions, Phase II focused on individual desktop applications (SPLASH) that could be linked through Peer-to-Peer technology to a network of community servers (PONDs) to form a global "POOL" of scaleable learning object repositories. SPLASH is a peer-to-peer learning object repository application freely available for download from www.edusplash.net.

In 2002 CANARIE funded eduSource Canada, a consortium to build a common infrastructure for learning object repositories in Canada. EduSource was funded for 18 months, to end March 2004, and is comprised of five development clusters led by Netera. The SFU Surrey cluster, with its expertise in interoperability mechanisms and knowledge media is developing the eduSource Communications Language (ECL) based on the IMS Digital Repository Interoperability Specification. SFU Surrey also extended and revised SPLASH, and will build a number of demonstration gateways to other important object networks that use Z39.50 and OAI protocols. The Surrey team has other deliverables in Collaborative Object Evaluation and Semantic Web mechanisms for global and local ontological mappings.

Edutella

Edutella is a peer-to-peer service for the exchange of educational metadata. Edutella lives on top of the Semantic Web framework as a distributed query and search service. The Edutella project addresses the shortcomings of current P2P applications by building on the W3C[37] metadata standard RDF. The project is a multi-staged effort to scope, specify, architect, and implement an RDF-based metadata infrastructure for P2P-networks based on the recently announced JXTA framework. The initial Edutella services will be *Query Service* (standardized query and retrieval of RDF metadata), *Replication Service* (providing data persistence/availability and workload balancing while maintaining data integrity and consistency), *Mapping Service* (translating between different metadata vocabularies to enable interoperability between different peers), *Mediation Service* (defining views that join data from different meta-data sources and reconcile conflicting and overlapping information) and *Annotation Service* (annotating materials stored anywhere within the Edutella Network).

Comparison of These Efforts with LionShare

The following information provides a comparison of other P2P effort outside US Higher Education related to P2P development.

SPLASH

The SPLASH effort has some similar features to the LionShare Project. SPLASH is based on the same metadata structure (IEEE-LOM 1.0), is developed in the JAVA programming language, both projects are in OpenSource, and are built on similar value propositions. These value propositions are: 1) learning objects are best stored close to their creators and users and will be shared freely in communities of practice, and 2) a single easy-to-use interface is essential for accessing and managing learning object collections, both large and small. There are also a number of dissimilarities between eduSource and LionShare: 1) eduSource has chosen to develop their own peering protocols, 2) have no concept of a PeerServer, have no authentication/authorization strategy on their P2P network, 3) have no authentication/authorization model, 4) use *only* metadata, and 5) don't have the facilities to store objects in a persistent way. Regardless of the dissimilarities, there are significant advantages to joining forces on the LionShare proposal. The projects are very complementary in that each development team has concentrated on different research agendas and combined there is significant concentration of expertise and source code to share. The LionShare team is farther ahead with the authentication and authorization strategies, in addition to centralized+decentralized strategies that incorporate the PeerServer. The eduSource team has made significant progress with the

connecting P2P networks to centralized learning object repositories that will bring large amounts of resources together in a single search.

Edutella

It is hard to really evaluate the Edutella efforts because the project seems in the early conceptual stages. There is a fair amount of technical documentation about this effort but there is very little product to judge the efficacy of the conceptual approach. There have been no software releases to date. There are also technical concerns regarding the peering protocol that is being proposed by Edutella. Edutella has based their development on the JXTA peering protocol which yet remains proven to be scalable and robust. The Canadian abandoned JXTA after nearly two years of development due to issues of scalability and stability. One interesting approach that the Germans have taken is to base their metadata strategy on RDF. This is a departure from other learning object metadata efforts that, thus far, have been based on the IEEE-LOM V 1.0.

Deliverables

The LionShare application and infrastructure is the foundation for the other two major objectives of this project. Both P2P federation and the connecting of existing repositories require a stable and refined application. The following list of deliverables will be the core objectives for Part 1 of this project.

Application Development

LionShare Peer

- A stable application, installable on multiple platforms with ease
- Collection Management tools including automated metadata entry with user profiles and file header information
- Connectivity tools such as user-to-user chat and chat rooms for groups with similar areas of interest

LionShare PeerServer

- A complete overhaul from the ground up featuring an open source relational database backend
- A network auditing system for use by network administrators and researchers
- Web-based administration for configuration, monitoring, and research

LionShare Networking

- Optimization and stabilization of the existing network code
- Integration of the authentication code described with detail in Section 2 and 3 of this proposal
- Integration of the code outlined in Part 2, Federated P2P Networks
- Integration of the code outlined in Part 3, Connecting Learning Object Repositories

Documentation

- Introductory document outlining the concept of LionShare and the benefits of P2P in academia
- Technical document consisting of a detailed description of the LionShare architecture and documentation of the code base for developers

Open Source Initiative

- A Web site containing a brief description of the LionShare project, papers, downloads, help, documentation, and a developer's section featuring: javadocs, CVS instructions, bugzilla, and message boards

Personnel

Lead Researcher:

Michael J. Halm, Penn State University, 513 Rider Building, University Park, PA 16802, Tel 1-814-865-2159, email: mjh@psu.edu

Technical Lead:
Lead LionShare Programmer, Penn State University, Computer Building, University Park, PA 16802

Part 2
Federated P2P Networks

Background

The LionShare prototype is being developed as an enterprise-level, authenticated academic P2P network for the Penn State community as part of the VIUS project funded by the Mellon Foundation. The Penn State LionShare P2P network authentication strategy uses Penn State's enterprise-level Kerberos authentication services.

However, there is a demand and a need within academic circles to collaborate and share resources inter-institutionally in an authenticated, controlled, and secure manner. It is feasible that academic P2P networks, like LionShare, might be extended far beyond one institution if appropriate middleware solutions provide "bridge functionality" between institutions. This bridge would allow for authenticated users on one campus to be authorized to use resources on other campuses. Modifying an unauthenticated P2P network and extending it to support intra-institutional use is a relatively trivial task. But, having an inter-campus P2P network with authentication and authorization requires much more thought and effort. Authentication technologies are still largely institution specific, and do not support inter-operation with other organizations. Recently, the development of some exciting new technologies has made inter-institutional authenticated P2P networks an exciting possibility.

With the advent of Shibboleth, organizations use their existing infrastructure to authenticate users locally, and use the Shibboleth technology to sign and transport assertions about the user to the remote site. These assertions are then used to determine access privileges. This allows workgroups to share resources in a secure and controlled fashion.

Additionally, the development of application level Open Service Interface Definitions (OSID) by the Open Knowledge Initiative (OKI) provides a method to insulate the application from changes in the technology by providing an application interface. These two methods offer new opportunities to develop end-to-end services that leverage software development at multiple levels. It is the intention LionShare *Federated P2P Network* development team to collaborate in the early stages of development on the architecture for the LionShare and produce a joint White Paper (see Deliverable) outlining the architectural design strategy for the project to incorporate these two exciting new technologies.

The following sections provide a preliminary overview of how Shibboleth and OKI might be implemented in a complementary but further discussion with the entire development team is needed to finalize an appropriate architectural approach.

Shibboleth

The Shibboleth project provides an inter-institutional middleware infrastructure that enables an attribute-based access control framework. The origin campus authenticates the user. The campus then generates a signed authentication assertion containing a Shibboleth "handle". The digital signature indicates who created the assertion. The handle is a unique opaque reference to the user. The LionShare client can attach the authentication assertion to a request and send the resulting message to a target (perhaps on another campus). The target can use the handle to request attributes from the origin campus about the associated user, but cannot determine the user's identity by merely examining the handle. The origin campus, not the individual user, asserts the attributes. The target can then use the attribute values it has obtained to determine whether the user is authorized to perform the requested operation. A user's identity is one potential attribute, but there are many scenarios that work effectively and securely using other attributes (e.g. student in course PS163 @ PSU; member of the PSU community, etc).

Shibboleth relies on a PKI-based trust fabric to provide the required level of assurance between origin and target. User certificates are not required. Instead, certificates are issued to the Shibboleth entities, which use the matching private keys to sign assertions and to create authenticated SSL tunnels. Scaleable and manageable operation is obtained when institutions band together into federations. These federations manage and distribute the files used by members to determine and validate the identity of other speakers.

(Federations also often impose policy requirements on members, and promote interoperability by requiring agreement to standard definitions of commonly used attributes.) Internet2 has created the first federation — called InCommon.

Shibboleth provides policy mechanisms at both the origin and target sites to help manage the use of attributes. At the origin, Attribute Release Policies (ARP) determine which set of attributes is released to each individual target site. Both site-wide and user specific ARPs can affect the release set. An attribute resolver configuration defines from where the attribute values are obtained (the campus LDAP directory, one of the business system SQL databases, etc). At the target, Attribute Acceptance Policies (AAP) ensure that the origin is authorized to assert the received attribute assertions. For instance, the default set of AAPs would cause a target to discard an assertion from Brown University that the user is a member at Penn State. Targets also use resource manager policy to specify, for each action on each resource, the set of attributes that the user must supply.

Organizations employing Shibboleth do not have to change existing authentication mechanisms to exchange content with other Shibboleth-enabled organizations. The Shibboleth mechanism obviates the need for less desirable multiple digital identities, proxy servers, or IP based restrictions. Shibboleth uses the SAML message formats and protocols defined by the Security Services Working Group within OASIS.

The Shibboleth model presents some interesting challenges however when one considers how it might be used to federate P2P networks. While the concepts of Shibboleth (local authentication, privacy, federation, authorization) are powerful, they have been designed and implemented for Web-based services rather than as a generalized middleware authorization mechanism. This invites some interesting and challenging research and development opportunities for both P2P network architectures and distributed authentication and authorization strategies. Thus, the primary challenge of this portion of the LionShare project will be to leverage the Shibboleth architecture and concepts to support the centralized+decentralized P2P topology. For instance, currently the only way to present a handle to a target is by using a Web browser. The following discussion describes the major components of this effort to extend the Shibboleth design and implementation to operate in a client-server environment.

Project Description

There are a number of components that need to be developed to support a robust federated P2P environment. First, the LionShare Peer needs a mechanism to obtain a SAML authentication assertion without relying on Web browser protocols. Currently, the LionShare Peer uses the Kerberos authentication protocol to directly authenticate to another LionShare Peer. Second, a federated approach to access control needs to be developed that can be deployed in the centralized+decentralized P2P topology. Third, software solutions need to be integrated into the LionShare architecture (see *Figure 2.1. LionShare Topology*). See *Figure 3.1. LionShare Federation* for an overview of the P2P federation approach. Finally, a pilot project needs to be developed to test the solution in a variety of enterprise environments. The following discussion provides more detail on each of these elements.

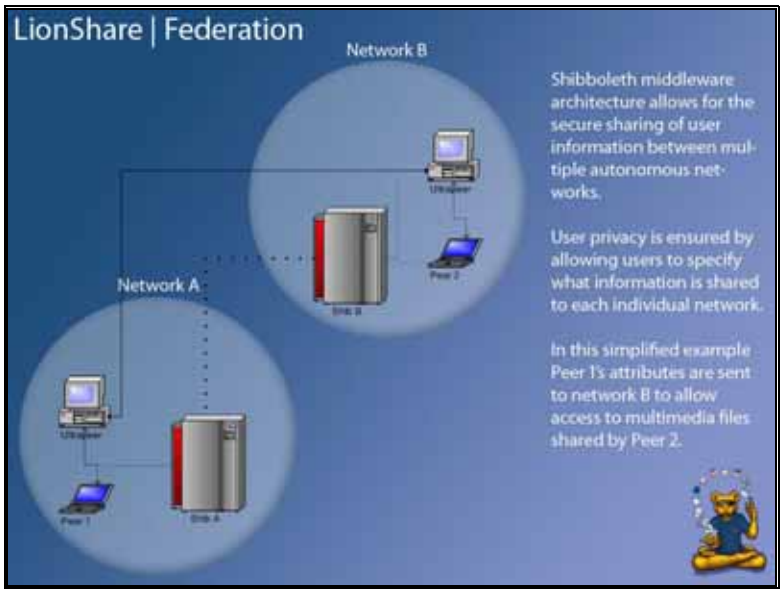


Figure 3.1LionShare Federation

Obtaining a Handle

The authentication assertion containing the handle has to be generated and signed by the origin campus. Consequently, a new network-based service will be developed to provide this functionality. The LionShare Peer will connect to this service and authenticate as the user. If authentication succeeds, the required signed assertion will be returned to the client. Some analysis will be required, to protect the system from attacks in which a third party somehow obtains a copy of this assertion.

Support for Various Forms of Authentication

A federated P2P must support a variety of popular authentication protocols at the level of the peer. The current version of the LionShare Peer is coded in Java and calls the Java Kerberos Libraries to connect to Kerberos services. To demonstrate P2P federation, a goal of this new effort is to provide campuses the flexibility to choose other authentication models. The protocol between the LionShare Peer and this new handle dispensing service should allow a user to choose from among a variety of authentication methods. (Obviously, though, the user will have to choose a method supported by their campus.) This protocol should use an existing protocol such as SASL to allow the LionShare Peer and this new service to dynamically negotiate a preferred authentication method. With this in mind, these are the forms of authentication that this effort intends to support:

- Lightweight Directory Protocol (LDAP)
- X-509-based Public Key Infrastructure (PKI)
- Kerberos
- Host User Name and Password

Authorization

Authorization is an essential component for the federation of P2P technologies. Over the last 18 months, there have been some successful efforts to define schema or languages that define generalized access control policy. Of particular relevance to Shibboleth (which uses the OASIS SAML protocols) is another OASIS effort. The XACML and SAML groups have worked to align their schema, so that both are using the same attribute assertions. Sun Microsystems has recently open-sourced their java-based XACML implementation.

If this can be integrated with the LionShare Peer (and PeerServer), it could provide a powerful and flexible access control solution.

However, XACML policies are described using a complex XML schema. A GUI client would have to be developed to allow "normal" users to create policies describing which users (i.e. those supplying which sets of attributes) are allowed access to the resources they manage.

Software Development Tasks

Adding Shibboleth support to LionShare, and providing a powerful access control mechanism, will require several distinct software development efforts:

1) Implement the Handle Dispensing Service (HDS).
2) Enhance the existing LionShare Peer to access the HDS. This should allow the user to specify their preferred authentication methods.
3) Enhance the LionShare Peer and Peer Server to support the SAML/Shibboleth protocol flows. This would include a) adding the authentication assertion to the initial request message, and b) when acting as a server, accepting and validating the authentication assertion, and using the handle to obtain attributes from the user's origin site.
4) Enhance the LionShare Peer and Peer Server to use the XACML engine to make an access control decision.
5) Develop a GUI for managing local policy (i.e. specifying who can do what to locally managed resources).

Ad-hoc Trust Fabrics

Shibboleth relies on a PKI-based trust fabric to identify and validate speakers and asserters. The initial LionShare pilots will leverage the (soon to exist) Internet2-managed InCommon Federation to manage the trust fabric. Campuses will be members of InCommon, and implement the Shibboleth services for their communities. Leveraging this infrastructure will vastly simplify the implementation of Federated P2P in LionShare. However, one of the strengths of P2P is its ad-hoc nature. P2P applications can oftentimes exist without relying on external infrastructure. A true Federated P2P application would allow a group to create and manage their own private federation, without relying on existing campus infrastructure. One obvious advantage of this approach is that the group could include members from organizations that do not operate the Shibboleth core services. However, freedom comes with a price—each of the members would have to install and manage their own Shibboleth installations, and they would have to create and manage their own federation. This is not impossible— there are already two national level federations in Europe. However, this process should probably be currently described as "requiring advanced level skills".

OKI

An OKI implementation of Shibboleth would require implementing both the OKI Authorization OSID and the Authentication OSID. While OKI is designed to support independent implementations of the Authorization and Authentication OSIDs (or indeed the implementation of only one of them), Shibboleth would require that these two OSIDs be implemented in a mutually dependent way. Maintaining two interdependent OSID implementations is an out of band agreement that must be enforced by the system integrator. The system integrator would do this by editing the app/implementation properties files to load only matched sets. The implementation for Authorization would have to be created so that it can use the authorization part of Shibboleth. The implementation of Authentication would have to be created to use the authentication part of Shibboleth.

OKI provides a means for sharing context information between OSID implementations through the Owner object. This is felicitous in the case of Shibboleth since OSID authentication proceeds with the assumption that the Agent's identity is contained in the Owner context. Additionally, it is important that what is saved in the Owner Context is serializable. This may mean that what is saved in the context is the information needed to get the Shibboleth objects rather than the objects themselves. The Shibboleth Target site sits behind the OSID implementations.

Project Organization: Federated P2P

The project leader for the Federated P2P effort will be Renee Shuey from Academic Service and Emerging Technology (ASET) group, part of Information Technology Services (ITS) at Penn State. Renee has played a key role in implementing the Shibboleth architecture at Penn State. She will have significant consulting support provided by Steven Carmody, Shibboleth Project Lead from Internet2/Brown University, as well as programming support for application software and Middleware development. James Leous from Academic Service and Emerging Technology (ASET) group, part of Information Technology Services (ITS) at Penn State, will provide additional technical support. In addition, liaisons will be assigned from other Internet2 Core Middleware areas. OKI architects and developers will also provide consulting support, architectural design expertise, writing and software development support towards this effort. These groups will provide consulting support, potential co-development opportunities, and a testing environment for the pilot project. Furthermore, this group will play a significant role in testing of the authentication and authorization of the LionShare P2P network, providing a heterogeneous enterprise authentication environment, as well as the testing of the authorization mechanism. A call for participation will be issued to the Internet 2 InCommon Federation, and the Internet2 mw-announce list. The group will have regularly scheduled conference calls, minutes, and a Web site supported by the Internet2 community.

Deliverables

There are a variety of deliverables for this portion of the project. These are the most notable:

- Federated P2P Architecture White Paper with Shibboleth and OKI participation
- Architectural Design for LionShare
- The Software Development Plan
- Peer and PeerServer software
- A Working Pilot
    - Federated P2P implemented in LionShare architecture demonstrating:
        - Resource discovery
        - Authentication
        - Authorization

Personnel

Lead Researhers:
Renee Shuey, Penn State University, 122L Computer Building, University Park, PA 16802, Tel 1-814-863-0005, email: rshuey@psu.edu

James Leous, Penn State University, 230 Computer Building, University Park, PA 16802, Tel. 1-814-863-7206, email: leous@psu.edu

Technical Lead:
Steven Carmody, Brown University, Box 1885, Providence, RI 02912-1885, Tel. 1-401-863-7580, email: Steven_Carmody@brown.edu

Part 3:
Connecting Learning Object Repositories

Overview

Several of the major findings mentioned in the VIUS report have to do with the importance of faculty and student having access to appropriate resources that fit their particular instructional, research, or outreach needs. Thus, connecting and extending the reach of P2P networks to ever-growing digital library and learning object repository efforts potentially could greatly enhance the attractiveness of using federated P2P networks. The significance of sharing learning objects and resources for education is well understood and supported by several standardization efforts including (IMS ARIADNE, EDNA[24] and eduSource Canada). The IEEE LOM[25] provides a basic means for describing learning resources with a standard set of metadata. The next step is to support interoperability mechanisms between LOR systems so the real sharing can occur. Two major interoperability specifications aiming to address the similar goals are IMS DRI and OKI. Although they are addressing the same issue they do it at different levels.

IMS DRI describes a number of high-level functions such as search-and-gather that LORs should support to interoperate with other LORs and tools. The IMS specification is a major advance, but remains to be interpreted and constructed by local repository holders. On the other hand OKI specifies API interfaces to support interoperability between learning systems, repositories, and other associated support systems and tools.

While both of these systems represent major advances, they leave much work to the local repository developers. For IMS, the developers must define and implement their architecture and protocols to achieve the DRI functionality. There are currently no freely available code libraries of implementations for implementers to customize for their LORs. OKI addresses interoperability between components within a learning system and does not address the interoperability between similar systems or between other initiatives such as NSDL[26]. There is a need to merge these two approaches for a truly interoperable distributed learning system at the global and local level. Fortunately, the eduSource Canada project has built an infrastructure that may bridge these gaps.

eduSourceCanada and Its Outcomes

EduSource Canada is a major initiative to implement a pan-Canadian LOR infrastructure, *Figure 3.2 eduSource Infrastructure*, with the goal of providing an open source framework of interoperable tools for scaffolding LORs in education and training. The infrastructure implements the IMS DRI specification to connect three types of user communities: Individual users using P2P tools (POOL network), centralized repositories constructed by organizations, and repositories of third-party metadata harvested by communities of practice. The eduSource infrastructure puts emphasis on its ability to also provide userEducs with access to the resources from other major projects and initiatives such as NSDL (using OAI protocol) and standard library systems (via Z39.50 protocol).

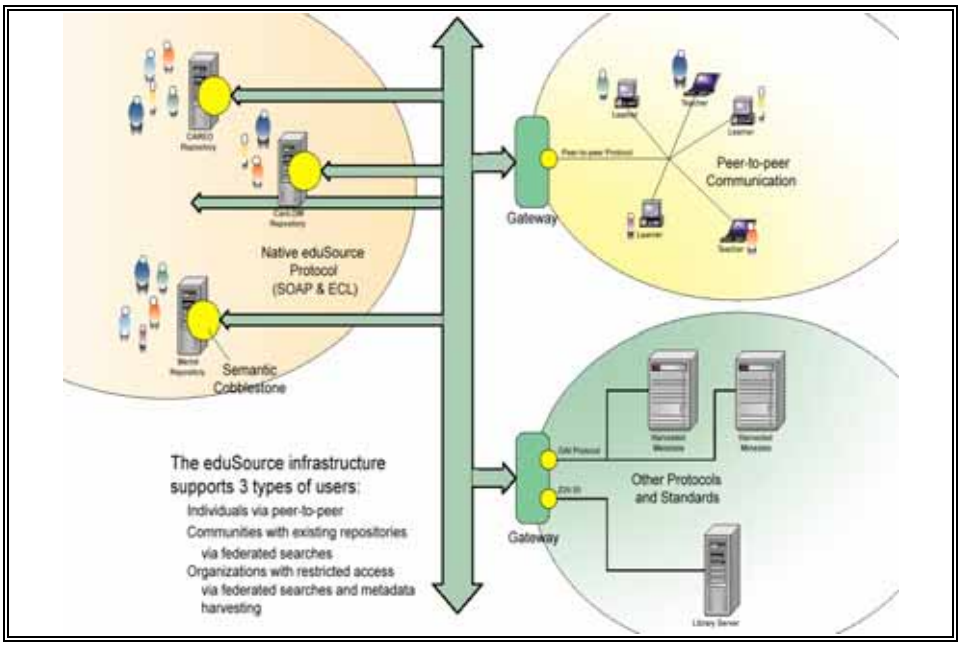


Figure 3.2 – eduSource Infrastructure

The eduSource Canada project ends in March 2004 with three major technical contributions to repository interoperability:

- a protocol named eduSource Communication Language (ECL) implementing IMS DRI
- a connector for connecting repositories to eduSource network. The connector is a configurable middleware component providing API for enabling the repository to communicate using ECL
- a gateway providing an expandable mediator service between ECL and other protocol. By the end of the project the transformation between ECL and OAI will be fully implemented. The gateway provides a generic architecture for plugging in other translators and within eduSource we will investigate the feasibility of transformation between ECL and Z39.50

Clearly, there is an opportunity to use the nascent infrastructure of eduSource Canada to help bridge between the significant communities forming around OKI infrastructure, IMS DRI protocol and OAI enabled systems such as NSDL. *Figure 3.3. LionShare Fixed Repositories* describes the basic functionality that this portion of the project will address.

Goals and Objectives

There are two main goals of this research proposal:

- To specify the activities and deliverables for the participation of the SFU team in the Penn State Proposal to link Peer-to-Peer LOR communities
- To specify the work plan for a similar venture with the MIT OKI proposal

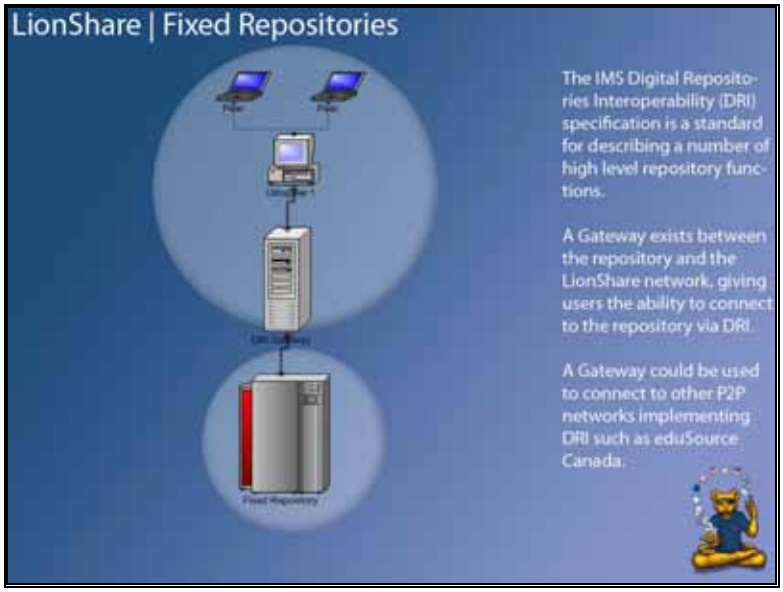


Figure 3.3. LionShare Fixed Repositories

Research Objectives

- To examine and implement a solution to achieve interoperability between Internet 2 based learning networks and eduSource. This includes:
    - Integrating Internet-2 Shibboleth Architecture with eduSource infrastructure
    - Providing bridges between eduSource and major eLearning network initiatives in Internet-2 space, such as LionShare. This involves defining the mapping between the protocols and expanding the eduSource gateway mechanism with appropriate mappings at transformation, communication, metadata, and vocabulary levels.
- To design and implement the connection between IMS DRI protocols as implemented in eduSource infrastructure and the OKI infrastructure, and if possible, the Z39.50 library communication protocol.

Research Outcomes

As eduSource infrastructure will not have a sophisticated access control mechanism the implementation of Shibboleth architecture within eduSource infrastructure will provide eduSource with open and sophisticated way of managing and executing access rights. More significantly, via this implementation the eduSource will become a first class citizen in the eLearning network under development in the US. It will be also a pilot implementation coupling IMS DRI with Shibboleth providing guidance for further development of the IMS DRI specification.

By achieving the second objective we will bridge two major interoperability initiatives, providing users of OKI compatible systems with access to distributed IMS DRI based systems, including the access to connected NSDL and other OAI based resources. For IMS DRI systems this will expand their repository-oriented interoperability with a set of interoperable learning tools via OKI infrastructure.

These achievements will directly benefit users both on the eduSource and Internet-2 sides. For example, LionShare users will gain smooth and secure access to resources of eduSource, which includes utilization of eduSource gateway connection to OAI-accessible NSDL resources for the LionShare users. EduSource users and connected content providers will benefit from the sophisticated Shibboleth architecture and will be able to access resources in other Internet-2 e-learning networks.

OKI and ECL

OKI has delivered a number of OSIDs that have direct application for the Connection Learning Object Repositories portion of this proposal. In particular, the Filing and Digital Repository OSIDs are of particular importance. The team envisions that these OSID will be of particular importance for bridging the connectors and gateways on the ECL as well as being useful in the Canadian and LionShare P2P application code on both ends of the network. OKI consultant will provide architectural design consulting, writing support for White Paper, and software development support for implementations.

Deliverables

- White Paper on architecture and strategies for connecting P2P networks to one another and to central repositories
- Software development plan
- Extension of the eduSource Communication Language (ECL) protocol with Shibboleth
- Shibboleth-enabled eduSource Connector (ShEC) and Shibboleth-enabled eduSource Gateway (ShEG)
- Guidelines for connecting repositories and tools using ShEC and ShEG
- Implementation of the protocol mapping for connecting peer-to-peer POOL network and one stand-alone repository using ShEC and ShEG into LionShare system.
- ShEG for OKI: Implementation of the protocol mappings into ShEG for connecting eduSource and OKI.

About SFU LOR Research Program

Within eduSource Canada, repository interoperability is the responsibility of the Simon Fraser University[27] research team of Dr. Griff Richards and Dr. Marek Hatala. This team is approaching four years of experience integrating knowledge management techniques, and emerging semantic Web technologies to address the needs of learning object repository communities. Together with their colleagues and students, this team has completed projects in community building, peer-to-peer learning objects, and is currently tackling the eduSource Canada interoperability needs. Their work on using semantic technology to resolve differential needs between global and local community metadata schemas has been presented at the International Semantic Web conferences, the WWW conference, and has been recognized with a number of peer research awards in Canada. The growing importance of this work examining the implications of the semantic Web for education has recently been recognized by Canada's NSERC funding agency for the next five years in a Learning Object Repository Research Network.

Personnel

Lead Researchers:
Dr Griff Richards, Simon Fraser University, 2400 Central City, 10153 King George Highway, Surrey, BC V3T 2W1, Canada, Tel 1-604-268-7432, email: griff@sfu.ca

Dr Marek Hatala, Simon Fraser University, 2400 Central City, 10153 King George Highway, Surrey, BC V3T 2W1, Canada, Tel 1-604-268-7431, email: mhatala@sfu.ca

Technical Lead:
Dr Marek Hatala, Simon Fraser University, 2400 Central City, 10153 King George Highway, Surrey, BC V3T 2W1, Canada, Tel 1-604-268-7431, email: mhatala@sfu.ca

Part 4:
Project Assessment

Assessment Approach and Objectives

The justification for this project is enhancement of the capability of faculty and students to share digital images, and assessment will answer the research questions specified earlier in this proposal as a means toward that end.

We intend to adopt an iterative assessment process similar to that used by many practitioners in the field of new product management. There is a considerable literature, especially in the marketing field, on effective user-oriented product innovation; the approach proposed here is consistent with well-established principles for merging needs assessment with design and delivery (1). The goal is to ensure that evaluation relates as closely and as realistically as possible to the critical options available for product development.

We believe that the process proposed here will be valuable because it overcomes the limitations of research alternatives that rely more heavily on predetermined questions and response choices. A more open, iterative approach has the advantage of maximizing the capability of informants (in this case, image users and members of the development community) to share their opinions, ideas, and knowledge. The theoretical advantages of this type of relatively less-structured data gathering have been recognized since at least the 1930s by social scientists and psychologists (2). Also, in practice, such less-directive methods have been widely accepted and employed over the past two decades because businesses in particular have found that they produce believable results at a reasonable cost. This is evidenced by the large and increasing numbers of corporate users who rely on methods such as focus groups for high-stakes advertising, manufacturing, pricing, and packaging decisions (3).

In short, systematic and thorough assessment will be conducted throughout the project, connected to and simultaneously with system development. The various assessment components will not be distinct, nor will they be sequential in terms of how they relate to development and implementation of LionShare. Nonetheless, because LionShare is being created as an enterprise-level, authenticated P2P network and because the project deliverables will include a working pilot, it might be helpful to think of the assessment component conceptually in two phases, which emphasize different goals and methodologies.

Exploration
The first phase primarily concerns early conversations with relatively small numbers of key users and developers. Demonstrations of the early LionShare prototype, produced by the recently completed Mellon/Penn State VIUS project, will be used to catalyze interviews and focus groups about whether and how P2P technologies can be made secure, scalable, and user-friendly. Information gathered in this stage will be useful for two purposes: first, to develop appropriate questions and instruments for future needs assessment, and second, to inform the many necessary detail-level decisions about LionShare features, interoperability, access rights control systems, and the like.

Refinement
The second assessment phase will be one of refinement. With improved questions and instruments, the team will seek input on the practicality, and desired features, of a P2P system (including priorities and likely tradeoffs) for a larger body of potential users. Paper and Web-based surveys will be emphasized in this aspect of assessment. Also, interviews, focus groups, protocol analyses, and monitoring of network usage can continue to provide valuable insights on LionShare features after they have been implemented, to improve the product based on user experiences and feedback.

To repeat, the above two-phase distinction is a bit of a simplification; it is somewhat artificial, since assessment will be ongoing and linked to product development.

In brief, the acute and often-justified criticism of needs assessment studies is that their results are too seldom used. Often findings are reported with no clear relationship to organizational change or program development (4). By contrast, this proposal commits to actions to integrate assessment findings to guide the development of LionShare.

Assessment Techniques

To answer our research questions, we intend to use a rich combination of assessment techniques.

- *Paper and Web-based surveys.* The VIUS project has found these to be useful, broad gauges of faculty, staff, and student perceptions, and persuasive when presenting findings to other audiences. Proposed members of the LionShare assessment team have access to Web survey software (*Remark*, from Principia), and extensive experience with survey development and administration (using the Web and other avenues). Web surveys will likely be especially useful for the assessment of a peer-to-peer system involving multiple institutions.

- *Intensive interviews and focus groups*. Members of the VIUS working team found both of these to be extremely informative. These results are less likely to embody the preconceived ideas of the team, and more likely than surveys to lead to surprising insights, and to bring unanticipated opportunities or hidden problems to the surface.

- *Protocol analysis.* This technique would perhaps be better labeled something like "think-aloud observational analysis". In any case, the tool was used during the VIUS project. It involves close, structured observation of users solving assigned problems. This research tool helps identify strengths and weaknesses in the design of an interface, especially as the design relates to the mental models and assumptions that users bring to a problem.

- *Network Usage.* The measurement of the activity of the P2P network can provide valuable information on the success of this initiative. The LionShare PeerServer has the capability of monitoring network activities and providing statistic on use. Additional functionality will be built into the final version to collect and analyze usage statistics.

- *Institutional Data Exchanges.* There are a number of opportunities for data gathering and assessment within forums conducted by Internet2 and Industry Canada. Internet2 has two yearly meetings, as does Industry Canada. It is the intention of the assessment team to hold feedback and working group discussions as part of these efforts. Interviews, focus groups and surveys can be used as a mechanism to collect information, face-to-face at these events. Requirements and methodology for gathering research findings will be discussed.

- *User field studies.* It is anticipated that individuals from a dozen different institutions in North America will connect to the LionShare P2P test-bed (that is, from among universities participating in the Internet2 Middleware Initiative Shibboleth project and the Industry Canada learning object repository development). Usability studies will be conducted with these individuals to explore, in particular, ways in which the application can be enhanced to promote adoption by the broader scholarly community. Those usability studies will probably be conducted for the most part by telephone.

Deliverables

Because assessment will be threaded throughout the LionShare project, deliverables from the assessment component will be reflected as an implicit, invisible part of the work products (system architecture, authentication, interoperability among federated P2P networks) of other aspects of the project. Nonetheless, LionShare will also include specific and explicit deliverables from the assessment dimension of the project.

- Papers and presentations — LionShare is intended to be useful to higher education colleagues working toward new models for sharing of personal and institutional virtual repositories. Therefore, as was true of the earlier Mellon/Penn State VIUS initiative, LionShare assessment results will be submitted for publication in appropriate journals and for presentation at appropriate professional conferences. Proposals for conference presentations will include, but not be limited to, Internet2 and Industry Canada yearly meetings.

- Mid-term and final reports —Assessment findings will be included in mid-term and final reports to the Mellon Foundation by the LionShare team.

Personnel

Lead Researcher:
Dr. Michael Dooris, Penn State University, Director for Planning Research and Assessment, 405 Old Main, University Park, PA 16802 Tel 1-814-863-8721, email: mjd1@psu.edu

Partners

About Penn State

Penn State's will use a team approach that will involve several different units to collaborate on the development of the *LionShare: Connecting and Extending P2P Networks* project. These include the Information Technology Services (ITS), University Libraries, and the Office of Planning and Institutional Assessment. Within ITS Teaching and Learning with Technology (TLT) and Academic Services and Emerging Technologies groups will plan an active role in the project. These groups will be responsible for software development of the LionShare P2P network and Authorization and Authentication implementations.

About Internet2

Internet2 is a consortium being led by 202 universities working in partnership with industry and government to develop and deploy advanced network applications and technologies, accelerating the creation of tomorrow's Internet. Internet2 is recreating the partnership among academia, industry, and government that fostered today's Internet in its infancy. The primary goals of Internet2 are to:

- Create a leading edge network capability for the national research community
- Enable revolutionary Internet applications
- Ensure the rapid transfer of new network services and applications to the broader Internet community.

Within the Internet2 organization two particular working groups will be participating in the project the P2P Working Group and the Shibboleth Project group. Information about these groups follows.

Internet2 Peer-to-Peer Working Group Mission

- To provide a forum for reporting on recent occurrences and future trends within the peer-to-peer and distributed computing space. The forum may occur at the regular member meetings, at Joint Techs meetings, or at specific workshops designated for the topic.
- To be a clearinghouse for collaborative opportunities within the higher education community and between that community and corporate entities as new peer-to-peer and distributed computing applications and tools are investigated.
- To provide best practices documents for both resource management as well as innovative uses of peer-to-peer technologies.
- To provide a central repository for resources and documents related to all aspects of peer-to-peer computing

Internet2 Shibboleth Project

Shibboleth, a project of Internet2/MACE, is developing architectures, policy structures, practical technologies, and an open source implementation to support inter-institutional sharing of Web resources subject to access controls. In addition, Shibboleth will develop a policy framework that will allow inter-operation within the higher education community.

About OKI

The Open Knowledge Initiative™ (O.K.I.) is defining an open and extensible architecture for learning technology specifically targeted to the needs of the higher education community. OKI provides detailed specifications for interfaces among components of a learning management environment, and open source examples of how these interfaces work. The OKI architecture is intended to be used both by commercial product vendors and by higher education product developers. It provides a stable scalable base that supports the flexibility needed by higher education as learning technology is increasingly integrated into the education process.

About eduSource Canada

The eduSource Canada project has created a test bed of linked and interoperable learning object repositories across Canada and provides leadership in the ongoing development of the associated tools, protocols, and practices that will support such an infrastructure. In 2002 Canarie funded eduSource Canada, a consortium to build a common infrastructure for learning object repositories in Canada. EduSource was funded for 18 months, to end March 2004, and is comprised of five development clusters led by Netera. The Simon Fraser University (SFU) Surrey cluster, with its expertise in interoperability mechanisms and knowledge media I developing the eduSource Communications Language (ECL) based on the IMS Digital Repository Interoperability Specification. SFU Surrey also extended and revised SPLASH, and will build a number of demonstration gateways to other important object networks that use Z39.50 and OAI protocols. The Surrey team has other deliverables in Collaborative Object Evaluation and Semantic Web mechanisms for global and local ontological mappings.

Section 6 – Structure and Organization of Project Team

The following discussion describes the project management and team structure for the LionShare: Extending and Connecting P2P Networks.

Project Management

The project will be managed and directed by a project steering team consisting of **J. Gary Augustson**, Vice Provost for Information Technology Services; **John Harwood**, Senior Director, Teaching and Learning with Technology, Information Technology Services; **Kevin Morooney**, Senior Director, Academic Services and Emerging Technology, Information Technology Services; **Bonnie Mac Ewan**, Assistant Dean for Collections, University Libraries; **Henry Pisciotta**, Arts & Architecture Librarian, University Libraries; **Michael Dooris**, Director for Planning, Research and Assessment; and **Michael J. Halm**, Senior Strategist, Teaching and Learning with Technology and Project Coordinator. The team will meet quarterly throughout the course of the project and provide project leadership and supervision. Michael J. Halm will be the project coordinator and manage the day-to-day project teams and supervise the software development and assessment activities. Lead researchers on the project will be Michael J. Halm, Renee Shuey, Marek Hatala and Griff Richards and Michael Dooris.

Application and Middleware Development Teams

There will be several development teams working on various deliverables. Each development team will have a team lead. Team leads will organize the individual work plans for their team and keep the team on task for completion of the team deliverables. Team lead for the LionShare team will be Mike Halm at Penn State. Team lead for the Federated P2P effort will be Renee Shuey at Penn State. Team leads for the Learning Object Network activities will be Griff Richards and Marek Hatala at Simon Fraser University in Canada. The size of the development teams will be based on the effort involved in producing the deliverables specified by each group. Team leads have provided detailed budgets for each teams' development activities. Development team leads will meet initially to kick off the project, and will maintain contact through regularly scheduled conference calls and occasional site visits when appropriate.

Assessment Team

Assessment has been a key component of the VIUS project that has led to the creation of LionShare and should continue to play an important role in the future development of the project. Usability and functional requirements need to be continuously monitored to assure that the project meets the goals and objectives set out in this document. An assessment specialist will be involved early to ensure the appropriate assessment questions are developed to answer the research questions developed for the project and to organize the assessment activities. Feedback at key points in the development cycle will provide input necessary to make adjustments in the design and implementation of the network and associated software components. To accomplish these tasks a part-time assessment specialist will be needed for the length of the project and will coordinate and direct the assessment activities. Michael Dooris, Director for Planning, Research and Assessment will be the Lead Researcher for this initiative.

Communications Team

This project will have an education component to help disseminate the findings of the project, as well as, educate our students, faculty, and staff about the legal and ethical issues of sharing files in a responsible way. LionShare provides an environment where files can be shared responsibly. Communications products will be distributed through a project Web site, presentations at technical and academic conferences, through articles published in appropriate public journals and by using the Internet 2 and Canadian eduSource communities. Regular news information will be syndicated through an RSS channel to appropriate organizations. The communication team will have a part–time technical writer assigned to develop communication products, Web resources and user education materials.

Project Planning

The project will be guided by several key planning documents: a software development plan, a communications plan, an assessment plan and White Papers that will prepared by Teams 1-3. Contingent on funding, team leaders will be required to develop a total project plan with benchmarks and deliverables for each of the these functional areas of the project. The software development plan will incorporate and organize the three areas of work so that workgroups can integrate the development effort of the other development teams.

Staffing

The breadth of this project requires a team approach in order to provide expertise in each area. We have assembled a Project Research Team composed of four Penn State professionals and two Canadian academic experts.

To emphasize this team approach, we have chosen the term "Project Coordinator" rather than the more customary "Principal Investigator" for the administration of the project. Five additional researchers will take responsibility for the portions of the project most suited to their specialties. Résumés for the principal researchers are attached as Appendix III.

- Project Coordinator and Researcher for P2P Application Development (*Michael J. Halm*, Senior Strategist, Teaching and Learning with Technology). In consultation with the principal researchers, builds and maintains the project plan, hires and supervises project staff, maintains the project schedule, and communicates with the funding agency. Works closely with each of the principal researchers to coordinate their efforts. Directly oversees the survey of existing image resources and benchmarking communications with other institutions. Oversees the design and implementation of all measures of the LionShare P2P applications development, particularly those portions involving LionShare Peer application features, PeerServer features, Internet2 Working Groups relations, and federation code integration.

- Principal Researcher for Federated P2P Development (*Renee Shuey*, Senior Systems Analyst, Academic Services and Emerging Technologies, Information Technology Services). Oversees software design and implementation for the working pilot, as well as the liaison with P2P application development team and Internet2 Middleware Working Groups.

- Principal Researchers for Repositories Interoperability (*Griff Richards* and *Marek Hatala*, EduSource, Simon Fraser University, University). Oversees software design and implementation for the P2P gateway and connector technologies for access to digital repositories, including testing and coordination.

- Principal Researcher for Assessment (*Michael Dooris*, Director of Planning Research and Assessment, Center for Quality and Planning). Oversees the design and implementation of the needs assessment work, particularly those portions involving surveys, interviews, focus groups, and nominal group techniques.

- The Research Programmer/Analysts will evaluate existing software tools, apply and document those selected, and write and document any original programs required for the creation of authentication logs and for development and implementation of the prototype. (Programmer/Analysts will be supervised by appropriate development teams.)

Funding requested for the above positions will be used to obtain temporary help to offset their current duties. The positions below will be hired or will be recruited as volunteers for the project period or for the length of testing required to provide quality assurance. Funding for these positions is requested.

A Project Associate will provide support for, and conduct portions of, the user studies. Skills in questionnaire development, focus group work, nominal group techniques, protocol analysis, and basic statistical analysis will be hired or developed through training at the outset of the project. In addition, this position will provide administrative support for the length of the project.

Focus Group and Interviews Participants will be needed to test and provide feedback during the development cycle. These will be offered a small monetary award for participation in these activities.

Production and Clerical Support will be hired on an hourly basis for interview transcription, scanning, data entry, document production, and related activities. This category of support may include use of Penn State's Statistical Consulting Center, which provides statistical consulting or production work on an hourly fee scale.

Consultants will be used to supplement the knowledge base for the project as needed. We expect that consulting services will be especially useful for advanced help with Shibboleth and OKI implementation.

Independent Testers will review the various software and technologies during the development process. We anticipate active participation from the Internet2 community as part of the software development cycle. These volunteer reviewers will be solicited from among qualified colleagues at other institutions.

Project Time-line

The plan of work may be conceptualized as four distinct components: P2P Application Development, Federated P2P Development, Repository Interoperability Development, and Assessment efforts. Projected timeframes for these phases are described in Table 5-1, *Overall Project Component Timeline.*

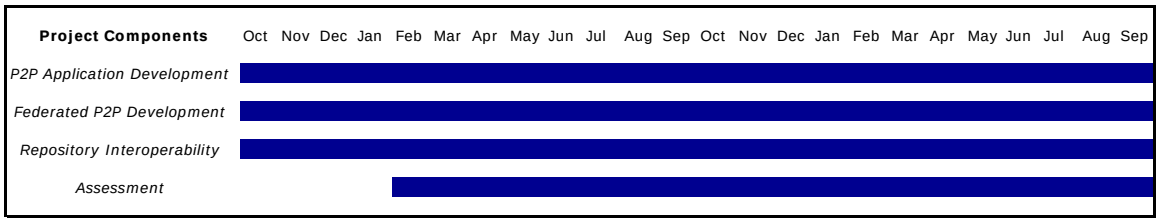
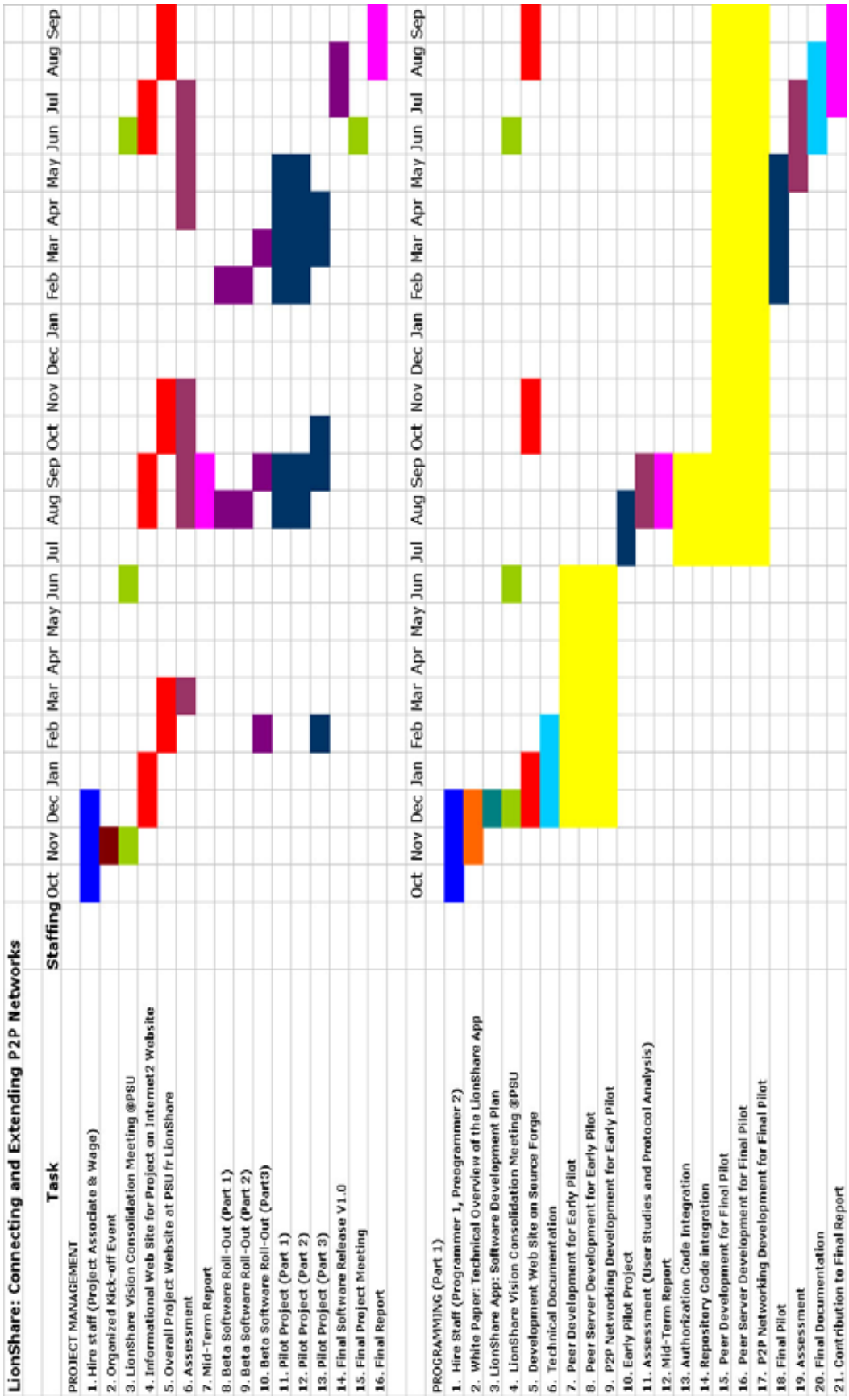
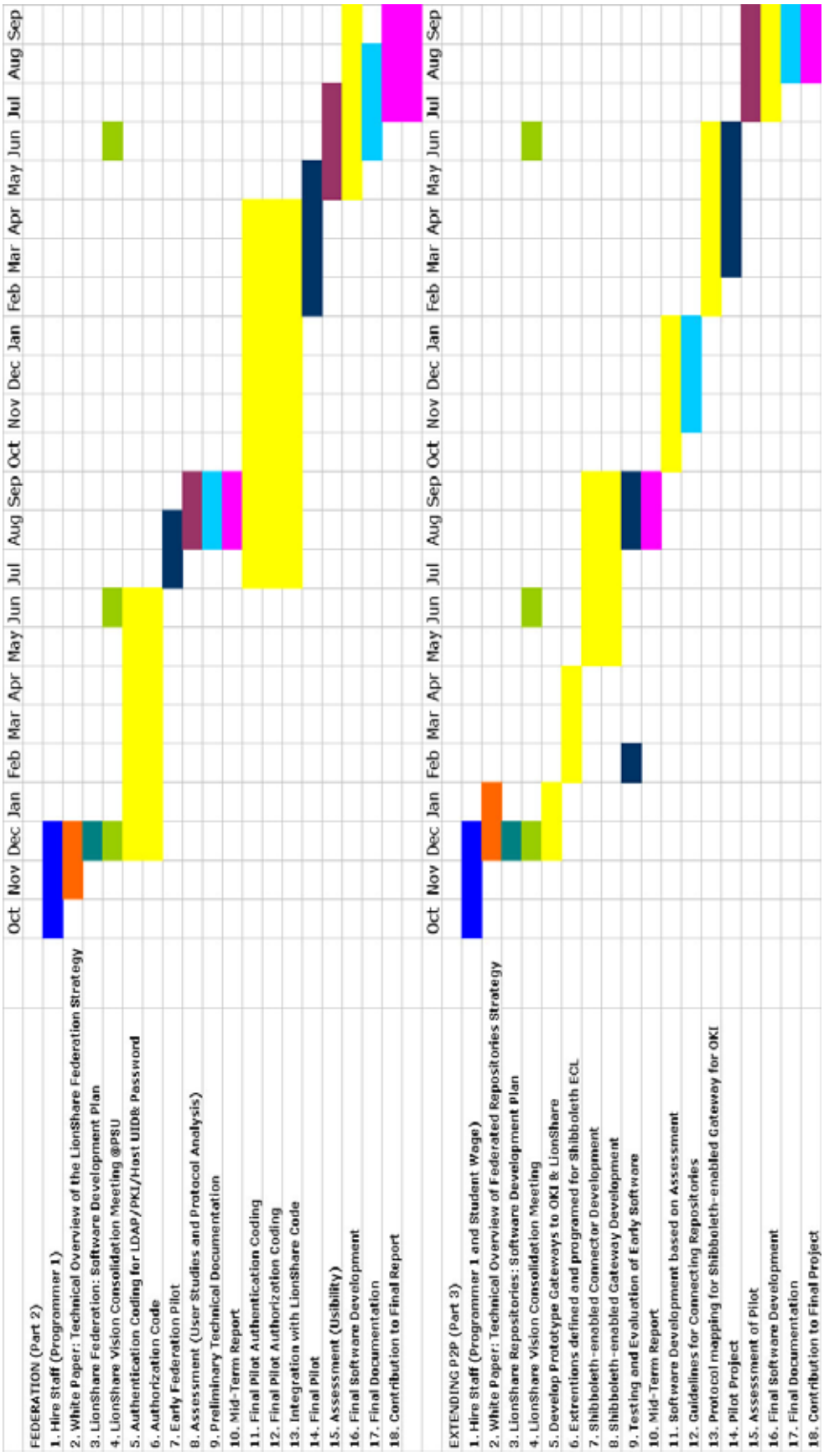


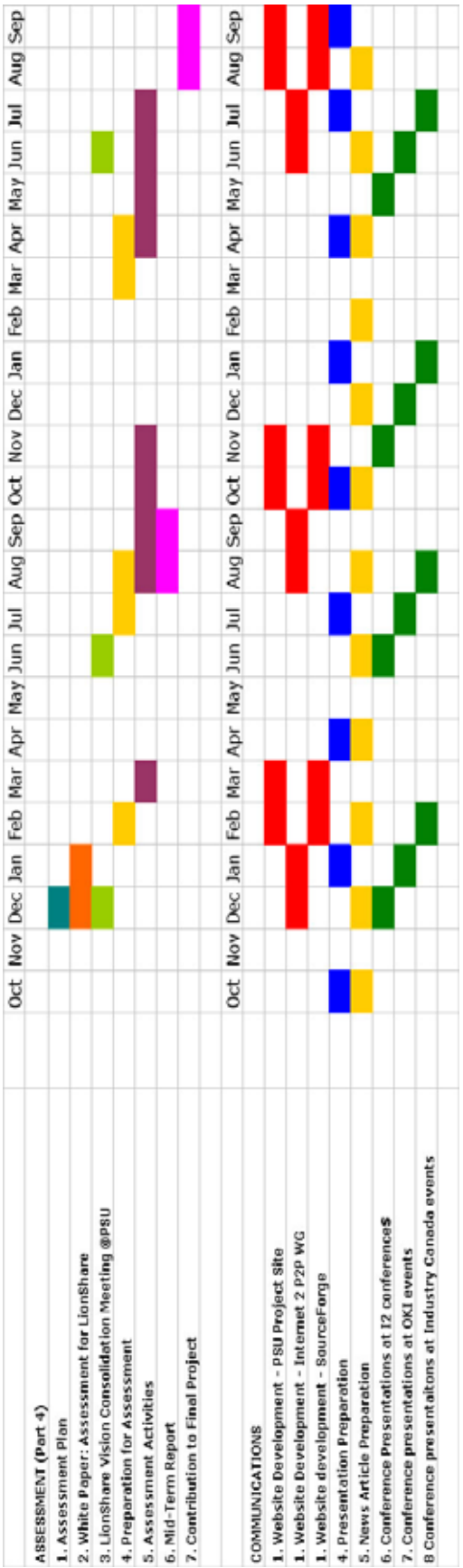
Table 5-1. Overall Project Component Timeline

Work Plan

The project period is planned to begin in October of 2003. For a more detailed schedule of tasks see Table 5.3. Project Plan – Project Schedule By Month.


LionShare: Connecting and Extending P2P Networks
Task
Staffing
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep
PROJECT MANAGEMENT
1. Hire staff (Project Associate & Wage)
2. Organized Kick-off Event
3. LionShare Vision Consolidation Meeting @PSU
4. Informational Web Site for Project on Internet2 Website
5. Overall Project Website at PSU fr LionShare
6. Assessment
7. Mid-Term Report
8. Beta Software Roll-Out (Part 1)
9. Beta Software Roll-Out (Part 2)
10. Beta Software Roll-Out (Part3)
11. Pilot Project (Part 1)
12. Pilot Project (Part 2)
13. Pilot Project (Part 3)
14. Final Software Release V1.0
15. Final Project Meeting
16. Final Report
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep
PROGRAMMING (Part 1)
1. Hire Staff (Programmer 1, Preogrammer 2)
2. White Paper: Technical Overview of the LionShare App
3. LionShare App: Software Development Plan
4. LionShare Vision Consolidation Meeting @PSU
5. Development Web Site on Source Forge
6. Technical Documentation
7. Peer Development for Early Pilot
8. Peer Server Development for Early Pilot
9. P2P Networking Development for Early Pilot
10. Early Pilot Project
11. Assessment (User Studies and Protocol Analysis)
12. Mid-Term Report
13. Authorization Code Integration
14. Repository Code integration
15. Peer Development for Final Pilot
16. Peer Server Development for Final Pilot
17. P2P Networking Development for Final Pilot
18. Final Pilot
19. Assessment
20. Final Documentation
21. Contribution to Final Report


Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep
FEDERATION (Part 2)
1. Hire Staff (Programmer 1)
2. White Paper: Technical Overview of the LionShare Federation Strategy
3. LionShare Federation: Software Development Plan
4. LionShare Vision Consolidation Meeting @PSU
5. Authentication Coding for LDAP/PKI/Host UID& Password
6. Authorization Code
7. Early Federation Pilot
8. Assessment (User Studies and Protocol Analysis)
9. Preliminary Technical Documentation
10. Mid-Term Report
11. Final Pilot Authentication Coding
12. Final Pilot Authorization Coding
13. Integration with LionShare Code
14. Final Pilot
15. Assessment (Usibility)
16. Final Software Development
17. Final Documentation
18. Contribution to Final Report
Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep
EXTENDING P2P (Part 3)
1. Hire Staff (Programmer 1 and Student Wage)
2. White Paper: Technical Overview of Federated Repositories Strategy
3. LionShare Repositories: Software Development Plan
4. LionShare Vision Consolidation Meeting
5. Develop Prototype Gateways to OKI & LionShare
6. Extentions defined and programed for Shibboleth ECL
7. Shibboleth-enabled Connector Development
8. Shibboleth-enabled Gateway Development
9. Testing and Evaluation of Early Software
10. Mid-Term Report
11. Software Development based on Assessment
12. Guidelines for Connecting Repositories
13. Protocol mapping for Shibboleth-enabled Gateway for OKI
14. Pilot Project
15. Assessment of Pilot
16. Final Software Development
17. Final Documentation
18. Contribution to Final Project


Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep
ASSESSMENT (Part 4)
1. Assessment Plan
2. White Paper: Assessment for LionShare
3. LionShare Vision Consolidation Meeting @PSU
4. Preparation for Assessment
5. Assessment Activities
6. Mid-Term Report
7. Contribution to Final Project
COMMUNICATIONS
1. Website Development - PSU Project Site
1. Website Development - Internet 2 P2P WG
1. Website development - SourceForge
4. Presentation Preparation
5. News Article Preparation
6. Conference Presentations at I2 conferences
7. Conference presentations at OKI events
8 Conference presentaitons at Industry Canada events

Appendix A – References

[1] VIUS. (2001). *Visual User Image Study* [On-line]. Available at: http://www.libraries.psu.edu/crsweb/vius/.

[2] IMS. (2002). *Visual User Image Study: Interim Report* [On-line]. Available at: http://www.libraries.psu.edu/crsweb/vius/reports.htm
[3} Insight Software, Luna Imaging, Inc. [On-line]. Available at: http://www.lunaimaging.com/insight/index.html

[4] MDID, The Madison Digital Image Database. [On-line]. Information available at: http://cit.jmu.edu/mdidinfo/

[5] CONTENTdm, DiMeMa, Inc. [On-line]. Information available at: http://contentdm.com/

[6] Jennifer Steiner, Clifford Neuman, and Jeffrey I. Schiller, *Kerberos: An Authentication Service for Open Networks.* 1988

[7] Lightweight Directory Access Protocol (LDAP). [On-line]. Available at http://www.ietf.org

[8] Open Source Initiative. [On-line]. Available at: http://www.opensource.org/

[9] Concurrent Versions System(CVS). [On-line]. Information available at: http://www.cvshome.org/

[10] eduSource Canada. (2003) *eduSourceCanada: Canadian Network of Learning Object Repositories.* [On-line]. Available at: http://edusource.netera.ca/english/home_eng.html

[11] Minar, Nelson, (2001) *Distributed Systems Toplogies: Part 1.* [On-line]. Available at: http://www.edna.edu.au/

[12] GNU GPL, GNU General Public License, [On-line]. Available at http://www.gnu.org/copyleft/gpl.html

[13] Gnutella Developers Forum. [On-line]. Information available at: http://groups.yahoo.com/group/the_gdf/files/

[14] PHP Hypertext Processor. [On-line] http://www.php.net

[15] Internet2 Middleware Initiative. [On-line]. Available at: http://middleware.internet2.edu/

[16] Shibboleth, *Shibboleth Project.*[On-line]. Available at: http://shibboleth.internet2.edu/index.html

[17] MACE, *Middleware Architecture Committee for Education*. [On-line]. Information available at: http://middleware.internet2.edu/MACE/

[18] CANAIRE. Canadian advanced internet development organization. [On-line]. Information available at: http://www.canarie.ca/

[19] SPLASH. The SPLASH networks. [On-line]. Available at: http://www.edusplash.net

[20] IMS Digital Repository Interoperability Specification. [On-line]. Available at: http://www.imsproject.org/digitalrepositories/index.cfm

[21] Java Runtime Enviroment (JRE). [On-line]. Information available at: http://java.sun.com

[22] Instant Messaging (IM). [On-line]. Information available at: http://computer.howstuffworks.com/instant-messaging.htm

[23] Public Key Infastructure (PKI). [On-line]. Information available at: http://www.pki-page.org/
[24] EdNA. Education Network Australia. [On-line]. Available at: http://www.edna.edu.au/

[25] IEEE. (2002). IEEE P1484.12.1/D6.4 Draft Standard for Learning Object Metadata Document. [On-line]. Information available at http://ltsc.ieee.org/wg12/

[26] NSDL. National Science Digital Library. [On-line] Available at: http://nsdl.org/render.userLayoutRootNode.uP

[27] Simon Fraser University. [On-line]. Information available at: http://www.sfu.ca/index2.htm
[28] MERLOT. (2000). *Multimedia educational resource for learning and on-line teaching website* [On-line]. Available at: http://www.merlot.org/
[29] IMS. (2003). *IMS Content Packaging Best Practices and Implementation Guide, Information Model, and XML Binding Specification.* [On-line]. Available at: http://www.imsproject.org/content/packaging/index.cfm

[30] IMS. (2003) *IMS Digital Repository Core Functions Best Practices and Implementation Guide, Information Model, and XML Binding Specification.* [On-line]. Available at: http://www.imsproject.org/digitalrepositories/index.cfm

[31] IMS. (2003) *IMS Learning Design Best Practices Guide, Information Model, and Information Binding.* [On-line]. Available at: http://www.imsproject.org/learningdesign/index.cfm

[32] IMS. (2003). IMS Learner Information Package Accessibility for LIP: Best Practices Guide, Information Model, XML Binding and Uses Cases. [On-line]. Available at: http://www.imsproject.org/accessibility/index.cfm

[33] World Wide Web Consortium (W3C). (2000) Simple Object Access Protocols (SOAP) 1.1. [On-line]. Available at: http://www.w3.org/TR/SOAP/

[34] World Wide Web Consortium (W3C). (2003) *XQuery 1.0: An XML Query Language.* [On-line]. Available at: http://www.w3.org/TR/xquery/

[35] International Standard Maintenance Agency. (2001). *The Z39.50 Document.* [On-line]. Available at: http://lcweb.loc.gov/z3950/agency/document.htm

[36] Open Archives Initiative. *The Open archives Initiative Protocol for Metadata Harvesting, Information.* [On-line]. Available at: http://www.openarchives.org/OAI/openarchivesprotocol.html

[37] World Wide Web Consortium (W3C). [On-line]. Information available at: http://www.w3.org/

[38] OpenURL. (2000) *The OpenURL Syntax Description.* [On-line]. Available at: http://www.sfxit.com/openurl/openurl.html
[39] Open University of the Netherlands. [On-line]. Available at: http://www.ou.nl/info-alg-english-introduction/index.htm

[40] Educational Modeling Language. [On-line]. Available at: http://eml.ou.nl/eml-ou-nl.htm

[41] IMS. (2003) *IMS Question and Test Interoperability: Overview, ASI Best Practices Guide, ASI Information Model, and ASI XML Binding.* [On-line]. Available at: http://www.improject.org

[42] IMS. (2003) *IMS Simple Sequencing Best Practices and Implementation Guide, Information and Behavior Model, and Information Binding.* [On-line]. Available at: http://www.imsproject.org/simplesequencing/index.cfm

[43] The Limewire Open Source Project. [On-line]. Available at: http://www.limewire.org

Appendix B – Brief Lead Researchers Biographies

Michael J Halm

Michael J. Halm is a senior strategist and manager for the special project activities for the Teaching and Learning with Technology group of Penn State University's Information Technology Services (ITS). Mike also serves as the co-chair on the IMS Technical Board representing Penn State and the Committee for Institutional Cooperation (CIC).

His current research interests include electronic learning environments, learning communities, knowledge building tools, digital media distribution networks, digital libraries, metadata and learning object repositories, peer-to-peer networks, and e-learning standards.

Mr. Halm has been working in the area of multimedia and emerging technologies as they relate to their integration in the delivery of instruction and services. He brings a long-standing interest in emerging technologies such as digital video production and distribution—including streaming technologies, multimedia application development, metadata strategies for indexing resources, electronic learning environments, and content management systems.

During his tenure with Information Technology Services, Mr. Halm has held a leadership role in various technology initiatives including the creation of the C&IS Wagner Training Center, Faculty Multimedia Center, the development of the Alumni Kiosk System and the Penn State Electronic Application.

Academically, he as taught in the Department of Horticulture, Landscape Contracting Program, the Department of Architecture, and the School of Information Sciences and Technology at Penn State. He has also participated in a number of funded research projects including initiatives with funding from IBM, Apple, AT&T, Lucent, USDA, CIC, Mellon, the New Media Centers Consortium and others. Mike holds a Bachelor's degree in fine art from Northern Illinois (1973) and a Master's degree in architecture (2001) from Penn State that focused on virtual learning environments based on a studio learning model.

Renee Shuey

Renee Shuey is a systems engineer and group leader in the Emerging Technologies Group of Penn State Information Technology Services (ITS). Ms. Shuey has been employed by Penn State for 13 years, most of which has been spent with the Administrative Information Services unit of ITS.

Renee Shuey's interest in authentication and authorization began with her involvement in the Committee for Institutional Cooperation (CIC) Research Projects Group. Ms. Shuey and the Research Project Group devised a way to share library resources between several of the Big Ten member schools via the Distributed Computing Environment (DCE). As a result of this work, she became a leader in the integration of administrative computing within the Penn State Kerberos 5/DCE environment. Her areas of expertise include the creation of client libraries to do DCE remote procedure calls (RPC) on Windows Operating Systems, integration and installation of DCE enabled Web servers, and integration of Kerberos enabled Web proxy servers.

As a member of the Internet 2 Middleware Shibboleth Pilot team she has led Penn State's pilot project with the Penn State Department of Physics and North Carolina State University WebAssign team. This project, the first successful implementation of Shibboleth, allows Penn State students authenticating with Penn State's DCE environment to access Web resources at NC State in a secure, authorized manner. In the first year of production, this service has allowed over 1,800 students in first- and second-year physics courses to access WebAssign materials.

More recently, as a member of the ITS Emerging Technologies Group, Ms. Shuey has led efforts to investigate the possibility of establishing a Public Key Infrastructure which could serve all of Penn State.

Ms. Shuey holds an associate degree in specialized business, computer information systems from South Hills School of Business and Technology (State College, PA) and has participated in numerous IBM Learning Services courses.

Dr. Marek Hatala

Dr. Marek Hatala is an assistant professor at the School of Interactive Arts and Technology at the Simon Fraser University, Surrey, BC, Canada. Before this appointment he was an assistant professor at the Technical University of BC, Canada (2000?-2002), research fellow (1998-2000) at the Knowledge Media Institute at The Open University, UK and assistant professor (1988-1998) at the Technical University of Kosice, Slovakia. He has received his Ph.D. in cybernetics and artificial intelligence in 1997 from the Technical University of Kosice, Slovakia. His research interests are knowledge representation and management, organizational learning, adding semantics to the Web, and intelligent information retrieval. He has published three books, more than 20 papers (journals, in proceedings from international conferences) in the fields of knowledge-based systems, expert systems, image processing, user modeling, and Internet.

Dr Hatala is a lead researcher on the interoperability in the distributed and peer-to-peer networks, especially in the eLearning area. He has been involved in research and development projects in Canada, UK, and Slovakia for more than 13 years. He holds an NSERC Research Grant for studying *Ontologies as Conceptual Models for Corporate Repositories*. This four-year research and tool development aims to support the building and widespread use of knowledge-based repositories. The main goal of this research is to develop techniques and tools to facilitate the end user in the process of annotating the documents with ontology concepts. The secondary goals include the study of relationships between ontologies and (XML-based) repository structures; maintaining and updating conceptual models and repository structures; and document retrieval using ontologies. Dr Hatala is also a principal investigator in the LORNET Network – a five-year research network funded by NSERC Canada.

Dr Hatala is the technical lead for interoperability in the $8.5mil eduSource project creating a test-bed of linked learning object repositories. He also led the development of the heterogeneous peer-to-peer network in the $1.6 mil Portal for Online Objects in Learning (POOL) project funded by CANAIRE Inc. Within the eduSource project, the interoperability solution aims at the development of the infrastructure and protocols that aim at interconnecting peer-to-peer, centralized, and harvested repositories by implementation of IMS DRI specification.

In the past, Dr Hatala participated in several international projects funded by the European Union. In 1 mil ECU ENRICH project (Esprit Project 29015) he was involved in building a contextually enriched archive for organizational learning for British Aerospace and Siemens Automation & Drives. The developed framework used distributed ontologies to annotate best practices and other working documents. In 420k ECU ENCODE (Copernicus 940149) he was responsible for visualization and debugging components for knowledge modeling tools and for the system integration.

Dr Hatala publishes extensively in the area of knowledge management, ontologies, repositories, and repository network. In the last five years he has authored and co-authored over 15 peer-reviewed papers in the international journals, conference papers, and book chapters.

Dr. Griff Richards

Dr. Griff Richards is a project leader in the Technology Centre of the British Columbia Institute of Technology, Burnaby, B.C. Canada since 1987. He is also an adjunct professor in interactive arts and technology) at Simon Fraser University Surrey. Dr. Richards served for three years (2000-2002) as research integration officer for the TeleLearning network of Centres of Excellence at Simon Fraser and is currently director of research integration for the new NSERC Learning Object Repository Research Network (LORNet). He recently led the $1.6M CANARIE Portals for Online Objects in Learning (POOL) project to create the POOL, POND and SPLASH architecture (2002), and currently leads the $1M NewMIC/SFU efforts on eduSource Canada looking at interoperability issues for Canadian learning object infrastructure. He is just completing a three-year Social Studies and Humanities Research Council grant (with Tom Calvert and Carl Bereiter) to archive and disseminate the outcomes of the TL-NCE.

Dr. Richards has both Ph.D. (1994) and M.A. (1983) in educational technology from Concordia University in Montreal. Prior studies included a B.Sc(1974) in genetics and chemistry, and a B. Ed (1978) from the University of Alberta. He also studied computer science and languages (French, Spanish) at the University of Ottawa, and spent several years doing R &D and management in computer assisted learning and interactive video at Canada's National Research Council, the Canadian Forces Training System, ACCESS Network and Westerra Institute of Technology before migrating to Vancouver.

He has over 50 publications and, working in collaboration with Drs. Tom Calvert and Marek Hatala over the past four years, he has produced two special journal issues, two book chapters, 11 published papers and given over 20 invited presentations in the areas of research integration and learning objects repositories. His research and development work in educational systems has been recognized with local, national, and international awards. He was a leading collaborator in the creation of the new $7.5M LORNet Research Network.

An associate professor of distance education with Athabasca University, Dr. Richards teaches systems theory and instructional design online. Fluently bilingual (English and French), he is skilled at bringing together large collaborative projects and making them succeed.

Dr. Michael J. Dooris

Dr. Michael J. Dooris in his current position as director of planning research and assessment in the Office of Planning and Institutional Assessment at Penn State, provides analytic support to the Office of the President for university-level planning and decision making. Responsibilities include institutional research on a variety of topics such as instructional workload, faculty demographics and career development, strategic performance indicators, and student course-taking patterns.

Michael has served Penn State since 1981 in a variety of data analysis positions, including in the university budget office and academic affairs. Earlier professional experience includes survey statistician, U.S. Census Bureau, Washington, D.C.; and information systems consultant, Arthur Andersen & Co, Hartford, CT. As an active affiliate faculty member in Penn State's higher education graduate program, he typically advises eight to ten masters and doctoral students and teaches two courses —financial issues in higher education and organization theory in higher education.

Dr Dooris holds a Ph.D. (Penn State, 1992) in higher education, with a supporting field of business administration; MBA (University of Rhode Island, 1981) with a concentration in management science; BS (Penn State, 1975) in economics, with a concentration in mathematical economics. Mike held a CIC academic leadership fellow in 1997-98. Academic awards include the 1989 journal prize from *Planning for Higher Education* and the outstanding graduate student paper of 1988 from the Society for College and University Planning.

Publications have appeared in *The Review of Higher Education*, *Assessment Update*, *NACUBO Business Officer*, *Proceedings of the American Economic Association*, *Journal of Applied Behavioral Science*, *Planning for Higher Education*, *New Directions for Institutional Research*, and a variety of other edited volumes. Consultancies include workshops on higher education research policy (organized by the American Council on Education) in the Czech and Slovak Federal Republics, and on strategic planning at the University of Ghana in Legon, Ghana, West Africa. He is a member of advisory panel for *Student Learning Assessment: Options and Resources* (2003) published by the Middle States Commission on Higher Education.