Project Description: ITR/PE+IM:
EduCommons: A peer-to-peer (P2P) system for sharing, discovering, and using learning objects

Principal Investigator: David Wiley
Utah State University
wiley@cc.usu.edu
Co-principal Investigators: Mimi Recker
Utah State University
mimi.recker@usu.edu
Andrew Gibbons
Utah State University
gibbons@cc.usu.edu

PURPOSE, INTELLECTUAL MERIT, & INNOVATION

The purpose of the proposed research project is to improve teaching, learning, and research through the design, development, evaluation, and dissemination of a software system called EduCommons. The system will be comprised of three parts; (1) the Peer, a peer-to-peer application capable of searching for and sharing resources directly between users, (2) the Recommender, a system which will gather usage statistics and actively recommend resources (including other users) to system users, and (3) the Editor, a collection of tools and wizards that will allow non-technical users to employ resources located in the system to effectively support teaching, learning, and research. These three system components will combine to create a ubiquitous educational content infrastructure, allowing users to share, find, and use resources currently existing on others' personal computers (like syllabi, lecture notes, quizzes, essays, data sets, and research instruments) but currently inaccessible to them.

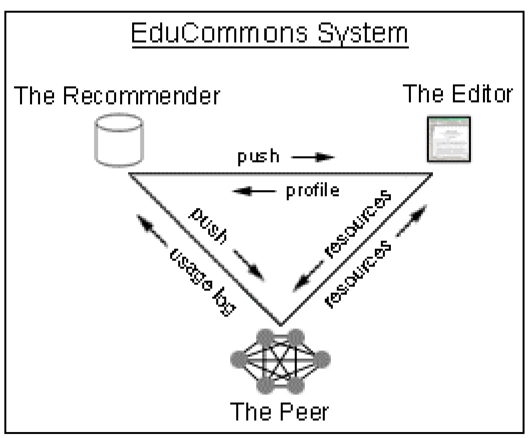

Figure 1. An overview of the EduCommons system.

While the Internet and its associated protocols (e.g., HTTP / the Web) have become the de facto resource access and distribution system of the new millennium, the Internet as we know it seems poised to disappear soon as the standard client-server model of computing gives way to a new computing paradigm known as “peer-to-peer.” In peer-to-peer (P2P) networks, no network node is privileged over any other (hence the term “peer”), and every peer is capable of communicating directly with every other. Applications such as Napster (2001), FreeNet (2001), and Gnutella (2001), all of which are distributed file-sharing systems, have provided the first popular examples of successful P2P architectures. These applications have also shown ways in which P2P architectures can be superior to traditional client-server models. For example, systems like Napster and FreeNet overcome the “Slashdot Effect” in which websites become decreasingly available as their popularity increases (because the webserver is incapable of serving the simultaneous demand). These systems do this by caching a copy of each resource locally when accessing it. These cached copies are then made available to the system at large. In this P2P architecture, resource availability **increases** with resource popularity, because each time a resource is accessed an additional copy is available to the system. By cutting out intermediary servers, the P2P model empowers users in a significant manner, which will make it a force to be reckoned with in the Internet of the future.

Simultaneously, the “learning object” model has emerged as instructional technology’s new paradigm. The learning object idea, in which educational content is stored in small chunks that are reusable in many learning environments, has gained such broad acceptance that the IEEE has formed the Learning Technology Standards Committee (with 20 working groups) to pursue the creation of standards for the description, interchange, and management of learning objects (LTSC, 2001). The authors of learning object systems praise their ability to reuse content in multiple learning environments, allowing for the creation of scalable learning environments with the potential to be significantly more efficient and adaptive than previous computer-based instructional systems (ADL, 2001; Gibbons, Nelson, & Richards, 2000; Wiley, 2000a). Within the academic field of instructional technology, learning objects have become a topic of intense research and publication, including the appearance of a recent dissertation (Wiley, 2000a) and book (Wiley, 2000b) dedicated to the topic, as well as numerous journal articles and book chapters (e.g., Bannan-Ritland, Dabbagh, & Murphy, 2000; Gibbons, Nelson, & Richards, 2000; Hannafin, Hill, & McCarthy, 2000; Merrill, 2000; Hodgins, 2000; Wason & Wiley, 2000; Recker & Wiley, in press).

Although they claim to focus on reuse of existing resources, current learning object systems require their content modules to be structured in semantically and visually specific ways; that is, these systems can only use and reuse specifically prepared and formatted content, precluding the reuse of the more than 16 terabytes of data existing on the publicly accessible Internet (Internet Newsroom, 1999). For systems whose main design goal is to increase the efficiency, effectiveness, and appeal of learning experiences through the reuse of existing resources, requiring all resources to be specially prepared is a significant weakness (Wiley, 2000a). This problem of facilitating the reuse of educational content is multiplied by the fact that the main storehouses of digital educational resources – the personal computers of students, educators, and researchers – are unavailable to other students, educators, and researchers.

This project proposes to explore ways in which the intersection of these two emerging technologies can catalyze significant advances in learning and research. We are specifically interested in improving teaching and learning through the use of these information technologies, and in learning more about instruction and learning themselves as we advance research in information technology. As more and more of teaching and training move online, understanding effective methods of creating and utilizing digital educational resources becomes increasingly critical. The proposed project will help develop knowledge fundamental to this crucial understanding.

AUDIENCE, BREADTH OF IMPACT, DIVERSITY, AND INTEGRATION OF RESEARCH AND EDUCATION

The proposed system will benefit any Internet user who seeks increased efficiency and efficacy in their educational quests, from pre-kindergarten children through seniors. This also includes students outside of the mainstream educational system (disadvantaged, disenfranchised, or home schooled students), not only in that they will gain access to a significantly larger amount of educational resources than are currently available to them, but through provision of a means for members of these communities to share resources directly with each other. Initially we will target two groups in information technology education. The first group is comprised of undergraduate information technology educators and their students at Utah State University and Marshall University (see attached letters of support). Because these universities have computer literacy requirements, all undergraduate students at these institutions could be impacted by the project (approximately **25,000 undergraduate students**). The second general group is comprised of K-12 students, their teachers, and college pre-service teachers who belong to either the Utah Education Network (UEN) or the Georgia Association for Information Technology (GAIT) with approximately 25,000 member educators, reaching over **880,000 K-12 students** (see attached letters of support). Because both Utah State University and Marshall University have extensive outreach programs they both serve considerable **rural populations** (Marshall University is located in West Virginia). The K-12 organizations also reach **ethnically diverse populations**, with over 45% of Georgia's K-12 student population coming from minority ethnic groups.

Additionally, the initial system development will be housed within the Department of Instructional Technology at Utah State University. The Department, currently ranked in the top five nationally, enrolls approximately 100 masters and doctoral students. These students comprise both in-service teachers in the state of Utah and people planning careers in instructional design and technology.

We propose to involve these audiences in several ways. First, by working in close collaboration with teachers and students in both undergraduate and K-12 settings, we will be able to gather valuable design guidance and feedback regarding user requirements and real world performance of the system. Second, by providing graduate employment for students at the Utah State University, several graduate students will gain valuable hands on research experience as they gain knowledge in their field. Teacher-researchers and their students within the initial target populations will also experience this same integration of research and educational experiences, which makes learning about information technology and the learning process itself exciting and vibrant.

Our design and development will follow the open source software model (see http://www.opensource.org/), in which source code is publicly distributed to the user community. In this way, other programmers can freely use, modify, improve, and re-distribute software much more quickly. Open source software formed and still comprises many basic building blocks of the Internet (Netcraft, 2001). These software initiatives evolved through a collaborative business model based on shared intellectual capital, as does the proposed system. They follow guidelines such as “release early, release often” and “given enough eyeballs, all bugs are shallow,” implying a thoroughly peer-reviewed rapid prototyping methodology. We plan to benefit from the additional support of open source programmers as well as contribute to the community by licensing software and documentation produced for the project under open source licenses such as the GPL (see http://www.gnu.org/copyleft/gpl.html) and OPL (See http://opencontent.org/openpub/). Making project documents and software open source **helps guarantee the broadest possible project impact**.

OBJECTIVES AND RESEARCH QUESTIONS

The objectives necessary to successfully complete the proposed project are grouped below in three large categories, called Content, Educational Context, and Learning Architecture. Relevant research questions immediately follow each objective. These objectives will be achieved through a series of iterative design cycles using a rapid prototyping design methodology. Ongoing formative evaluation results will serve to inform subsequent design phases.

<u>Content</u>

The **primary objective** of the content section of the proposal is to create a ubiquitous educational content infrastructure (UECI): a system in which a significant number of diverse educational resources can be shared and located for use and reuse to improve teaching, learning, and research. **Specific design and development sub-objectives** include:

1. Design and develop a peer-to-peer application and protocol that enables seamless retrieval of educational resources such as text, data, visual, and other available information in all subject areas to support teaching, learning, and research.

Several questions must be answered during the design and development of such an application. Is a pure peer-to-peer implementation necessary? Desirable? Which of the many peer-to-peer architectures or topologies is best suited to the application? Can the system piggyback on an existing protocol, such as HTTP, or is a new protocol necessary? As the WWW has become synonymous with ubiquity, how can peer-to-peer systems be made accessible through the WWW and existing browsers?

Many open source P2P implementations exist which can be used as starting points for this development (e.g., Gnutella, FreeNet, and OpenNap). Additionally, because these are open source / distributed development efforts, the design documents are publicly available as well. These projects have already struggled with issues such as architecture and scalability in non-educational domains. The existence of these resources will give the proposed project an excellent starting place. Additionally, the authors have secured seed funding in order to begin the immediate analysis and design of the system (see attached letter of support). Thanks to open source software, design documents, and seed funding, the proposed project will hit the ground running.

2. Design and develop adequate resource search / discovery mechanisms within the UECI.

While "metadata!" is the current battle cry, is metadata alone both a necessary and sufficient discovery-enabling mechanism for textual resources that can also be full-text indexed? (Metadata is descriptive information about a resource, such as the information on a card in a library card catalog. In the same way the card catalog allows users to find books without walking the entire library opening every book, metadata allow users to discover educational resources without directly examining each individual resource.) When metadata indexing is used, what is the optimal set of descriptors (e.g., author, title) and size, assuming that metadata adds value but is costly to create? What is the role of collaborative filtering or recommender systems in resource discovery? What type of supporting data would be needed to support such a system?

Because the IEEE's Learning Object Metadata Standard (LOM, 2001) for educational resources already exists, questions regarding metadata use can be narrowed from "what metadata fields are needed to support educational resource description?" to "which LOM fields should be used in this particular application?" Using a proper subset of the LOM will be important to insure future interoperability between the proposed system and other standards-based learning object systems. As for the role of recommender systems in resource discovery, mixed-method collaborative filtering approaches based on metadata and full-text indexing variants have been explored previously (Basu, Hirsh, & Cohen, 1998; Balbanovic & Shoham, 1997) and should be adaptable to the current educational application.

3. Facilitate the population of the UECI with a significant number of diverse educational resources by lowering the barriers to participation and establishing a sufficient incentive structure.

Questions related to this sub-objective include: How do we encourage users to contribute to and search through a content repository which initially has little content? How much content must be contributed before the system reaches an inflection point and begins to draw users? Why and how do people contribute resources to the public good? Do users expect some kind of reward, even if it is non-monetary (Rheingold, 1993)? How many users must contribute to the system (compared to the number that simply "leech" off the system) for it to succeed? What are the barriers to participation in such a system (privacy concerns, the complexity of the sharing process, intellectual property concerns, etc.) and how can they be overcome? How can expression of diverse viewpoints be encouraged within the system?

Recent research suggests that participation rates in peer-to-peer systems do not necessarily need to be high in order to succeed. For example, Oram (2000) found that only an average of 7% of users contribute to the Usenet News system. Yet, by most measures this system is extremely popular, with many millions of active users. In a recent study, Adar and Huberman (2000) analyzed traffic in a P2P file-sharing application called Gnutella, and showed that 70% of Gnutella users share no files whatsoever. This phenomenon of low contributor to consumer ratio has been dubbed the "tragedy of the commons" in that users may be tempted to free-ride off of the work of other users. Eventually, the system collapses as consumers overwhelm contributors. Shirky (2000) however, suggests that P2P applications are not subject to this paradox – simply because a digital good can be simultaneously accessed by many users. In economic terms, this is referred to non-rival use (Kollock, 1999). The new scenario in which nonrival digital resources are shared has been called the "cornucopia of the commons" (Bricklin, 2000).

Formative data regarding the Content objectives will be gathered by undertaking the following **evaluation objectives**:

A. To what extent does the system provide seamless retrieval of educational resources such as text, data, visual, and other available information in all subject areas?
B. To what extent does the system increase access to needed resources for populations currently being underserved?
C. To what extent are metadata taxonomies and user interfaces to them designed in such a way that novice users are capable of using them to find relevant educational resources?
D. To what extent does the system significantly increase the amount and quality of educational resources available online?
E. To what extent does it increase the diversity of educational resources available online?

<u>Educational Context</u>

Simply providing unlimited access to unmediated educational content will not solve all instructional problems. Otherwise, libraries would have provided the solutions long ago. Instead, we argue that effective access to and use of learning content must be also be augmented by contextual information. The **primary objective** of the educational context portion of the proposal is the creation of a 'context layer' for distributed educational content.

The notion of context includes any contextual (i.e., not content) information associated with a learning object. This includes such things as who authored a learning resource, why, how it was used in a learning situation, and by what kinds of students. It also includes issues of how people will be motivated to contribute to the content infrastructure. Contextual information is important because it can help users filter through large repositories to locate relevant content (Recker & Wiley, in press). Context can also be used to recommend unseen resources and like-minded individuals, for possible future collaboration. Finally, context can be used to help personalize content within new instructional structures.

Specific **design and development sub-objectives** related to educational context include the following.

4. Design and implement a 'context layer' for distributed educational content. This layer complements the content infrastructure, and captures how, why, and by-whom learning resources are contributed and used. This context information will be captured via the P2P application, and be sent to a centralized data center for further processing. As such, the approach advocates a value-added server-based service, and is not a pure P2P implementation.

Within this objective, the following research questions will be addressed: What kind of contextual data should be logged in order to facilitate the discovery and effective educational use of resources? Are implicit data such as overall object usage statistics more useful (Recker & Pitkow, 1996)? Alternatively, are explicit data such as users' perceived value of objects or user-provided annotations more useful? Would the cost of collecting explicit data outweigh its value? How much personal information about users (their "user profile") is necessary without comprising user privacy? How much is sufficient? How much will users be willing to share?

Many P2P applications focus on user privacy through absolute user anonymity (e.g., Publius, FreeNet, Gnutella). However, under current academic citation practices, if author names are not attached to resources the system will not provide a means for serious scholarship – it will only provide a means for plagiarism. In order to facilitate the proper citation of sources when resources in the system are used, it seems likely that users will have to voluntarily sacrifice some privacy when sharing resources (but not necessarily when searching for or using resources). Moreover, in an educational context, providing the author of a resource helps establish the authenticity and authority of the particular resource.

5. Using the data gathered, employ collaborative filtering algorithms to adaptively recommend (or "push") both additional relevant learning resources and like-minded people to users.

What collaborative filtering algorithms work best? Recker, Walker, and Wiley (2000) describe ways in which collaborative filtering systems can be used to recommend relevant educational resources to users, but are such recommendations valuable to users? In terms of privacy, will users allow themselves to be recommended to other users, creating spontaneous communities of interest? Do users value the recommendation of other users?

Walker (2001) reviews the collaborative filtering literature and its potential application to educational resource recommendation. Experience from our Altered Vista and

Instructional Architect projects (NSF-DUE-0085855), in which we are implementing systems for recommending educational resources will provide us with additional leverage toward completing the proposed project successfully.

6. Examine intellectual property rights within a distributed system for educational content.

How will students, educators, and researchers want to manage the “rights” to their content? Will system users simply release their content into the public domain? Will they elect to employ a form of “copyleft,” such as the Open Publication License or the OpenContent License (Wiley, 2000c)? Will rights management issues scare some users away from the system?

Through the OpenContent project we have gained valuable experience in working with individuals and commercial publishers on issues of freely sharing content while protecting the rights of authors (Wiley, 2000c). We will pair this experience with guidance from Raymond (1999) and Kollock (1999), who describe the economics of creating and distributing digital “public goods.”

Formative data regarding the Educational Context objectives will be gathered by undertaking the following **evaluation objectives**:

A. To what extent does contextual information improve the quality of searches performed in the system?
B. To what extent do system users find the recommendations of resources valuable? To what extent do system users find the recommendations of like-minded people valuable?
C. How do users feel about intellectual property and their privacy in the context of a distributed content sharing system? How important are these issues in order for the underlying system to succeed?

<u>Learning Architectures</u>

As stated above, providing unlimited access to unmediated educational content will not solve all instructional problems. Even a system that provides users with the ability to locate *relevant* learning resources is only half of an instructional solution. For this reason the content and educational context portions of the proposed study must be augmented by the consideration of learning architectures; in other words, the ways in which several learning objects can be combined into a coherent, instructionally valuable whole. While existing learning objects methodologies (Wiley, 2000a; Merrill, 1999) require that the resources be specially prepared and formatted, such preparation would likely present an unreasonably high barrier-to-entry for contributors to the proposed system. We therefore propose to explore methods of scaffolding the instructionally sound use of diverse learning objects by creating instructional support tools that will allow users to combine, organize, sequence, and add further context to educational resources they discover within the system. The **primary objective** of the learning architecture portion of the proposal is the creation of a set of instructional support tools that facilitate the instructional use of educational resources of arbitrary characteristics. Specific **design and development sub-objectives** related to learning architecture include the following.

7. Design and construct a family of instructional support tools to facilitate the instructional use of learning objects.

What are the relevant design dimensions of such tools? Is there a shared underlying architecture for the various instructional support tools? What types of learning activities must the tools be able to support? Does the requirement of supporting arbitrary learning objects limit the types of activities the tools can support? What level of understanding of the instructional design process can the tools assume?

Our current research suggests that a common architecture may underlie all instruction, from tutorials to high-end simulations (Gibbons, Nelson, & Richards, 2000). This common architectural understanding will be applied to the design of the tools. Merrill (2000) has described the manner in which current learning object-based instructional approaches relate to the learning process from a cognitive psychology perspective. Following Merrill we will explicitly ground the design of the tools in current research on learning. Additionally, in order to facilitate collaborative work in addition to (the traditional) single student computer-based training model, special attention will be paid to the collaborative process as outlined in Nelson's (1999) Collaborative Problem Solving model.

Formative data regarding the Learning Architectures objectives will be gathered by undertaking the following **evaluation objectives**:

A. To what extent are the instructional support tools diverse in their facilitation of types of learning? In the audiences they target?
B. To what extent are the needs of higher-order forms of learning satisfied by the instructional support tools?
C. How easily are untrained users capable of employing the instructional support tools?
D. To what extent do the instructional support tools facilitate the creation of high quality learning experiences?

DETAILED SYSTEM DESCRIPTION

Design, development, and evaluation will take place during all three years of the project. We will use a rapid-prototyping design methodology, in which frequent and ongoing evaluation involving user studies serves to inform subsequent design and development activities.

Technologies and System Architecture

The system will be initially implemented as three tools: two applications and a data center (server) component. These tools are described in more detail below.

1. **The Peer**. A peer-to-peer application for sharing and locating educational resources. This tool will be prototyped as a Win32 application, but will be ported to other environments as well (e.g., Mac, Linux). We will also explore WWW browser interfaces to the P2P network (e.g., Fproxy in Freenet). In addition to providing sharing and searching services, the Peer will contain indexing functionality that will automatically generate resource-describing metadata, and can optionally be enhanced through user-contributed metadata (for interoperability, this metadata will be a proper subset of the IEEE LTSC Learning Object Metadata set [LOM, 2001]). Search interfaces to the metadata will be designed following our previous research on metadata interfaces and design strategies, which suggests a phased search approach (Wason & Wiley, 2000; Wiley, 1999; Wason, 2001).

2. **The Recommender**. A collaborative filtering system consisting of a data center that gathers supporting data and an algorithm for the generation of recommendations based on supporting data. This paradigm is based on collecting and propagating word-of-mouth recommendations about the qualities of particular resources (Maltz & Ehrlich, 1995; Shardanand & Maes, 1995). Such algorithms have been used to implement recommender systems in a variety of domains, including recommending movies, research reports, and Usenet news articles (Resnick & Varian, 1997). These will both be implemented on a standard Linux server residing on the Utah State University's College of Education network. Usage data from Peer applications will be sent to the data center for storage and analysis, and individualized recommendations for resources will be transmitted to individual Peer applications. (Because of privacy concerns that may arise by system usage tracking of this type, use of the Recommender will be strictly optional.) The data center system will rely on a database of aggregated "contextual" information about resource use and reuse in the system. Communication between the applications and the database will use standard Internet protocols, such as HTTP. Algorithms underlying the recommender engine will rely upon a variety of data, including optional user profile information, resource metadata information, user-contributed comments, and resource usage. The latter algorithm will be based on a model derived in our prior research (Recker & Pitkow, 1996), which showed that object desirability is strongly correlated to recency and frequency of prior usage.

3. **The Editor**. A set of instructional support tools for combining, sequencing, and otherwise aggregating individual resources discovered in the system in order to support learning, teaching, and research. Example aggregations of resources include Goal-based Scenarios (Schank, Berman, & Macpherson, 1999), Open-ended Learning Environments (Hannafin, Land, & Oliver, 1999), and Computer-supported Intentional Learning Environments (or CSILE; Scardamalia & Bereiter, 1989). The Editor will be prototyped as a Win32 application, but will also be ported to other environments (e.g., Mac, Linux). This stand-alone application will eventually be integrated into the Peer application, providing "one-stop shopping" for educational resource sharing, discovery, and utilization.

We plan to benefit from the additional support of open source programmers as well as contribute to the community by licensing software and documentation produced for the project under open source licenses such as the GPL (see http://www.gnu.org/copyleft/gpl.html) and OPL (See http://opencontent.org/openpub/).

WORK PLAN

The three-year work plan is summarized in Table 1, and detailed in the following narrative. Parenthetical numbers in Table 1 reference the project objectives listed above.

Year 1	• Complete peer application design (1, 2) • Complete first peer application prototype (1, 2) • Document and release peer application source code following Open Source model (1, 2, 3) • Review and analyze existing collaborative filtering algorithms (5) • Derive data types to be gathered to support collaborative filtering recommendation (4, 5) • Design data center server software to harvest data supporting collaborative filtering (4) • Analysis, synthesis, and conceptual design of instructional support tools (7) • Create paper prototype of instructional support tools (7) • Begin computer-based prototype of instructional support tools (7) • User studies: metadata element, incentive structure, intellectual property, privacy, and HCI studies (3, 6) • Formative evaluation: expert review of tool prototypes, usability studies, observations and questionnaires
Year 2	• Coordinate Open Source development and improvement of peer application (1) • Implement collaborative filtering system (including data center) for adaptive, automated recommendation of educational resources (4, 5) • Document and release collaborative filtering implementation following Open Source model (4, 5) • Revise instructional support tools design based on formative evaluation (7) • Document and release instructional support tools source code following Open Source model (7) • Formative evaluation: expert review of tool prototypes, usability studies, observations and questionnaires
Year 3	• Coordinate Open Source development and improvement of peer application, collaborative filtering implementation, and instructional support tools (1, 4, 7) • Integrate peer application and instructional support tools (1, 7) • Summative evaluation: usability studies, analysis of system effectiveness, data center log file analysis, support for the formation of online communities • Dissemination: establish mirror sites and pursue licensing with commercial vendors

Table 1. A summary of project activities. Numbers in parentheses refer to project objectives listed above.

Year 1: Prototype Design and Development

The Peer. The project will begin where our current research leaves off: with seed funding from Utah State University's Vice President of Research's office, we are currently examining existing peer-to-peer protocols, architectures, and implementations in preparation for the proposed study (see attached letter of support). We will use this analysis as the starting point for the design of a peer-to-peer application (the Peer). A number of user studies must also be completed before design of the sharing and search strategy and interface can begin, including empirical studies of IEEE/LOM metadata element usage in existing LOM-compliant systems and user attitudes toward privacy and intellectual property. Following Wiley (1999), the primary search interface design philosophy is to reduce complexity of interaction for the user in two ways. First, metadata are subdivided into thematically similar groups. Second, sub-tools or wizards are provided for use when, and only when, the user needs them. The **result of the first year's work** on the Peer will be a functioning prototype capable of sharing, indexing, and searching for educational resources.

The Recommender. During the first year we will also conduct a thorough review and analyses of collaborative filtering algorithms as they apply to recommending educational resources, extending our current research (Recker, Walker, & Wiley, in press; Walker, 2001). The **result of the first year's work** on the Recommender will be the selection of an algorithm, the determination of what usage and optional user data to track, and a preliminary design for the data center database. User data tracking decisions will be based partially on user feedback.

The Editor. Finally, in the first year we will explore instructional support tools as scaffolds for educational resource use. An analysis of existing instructional design theories and practice (e.g., Reigeluth's [1999] overview) will be synthesized into several primitive instructional support tools. The **result of the first year's work** on the Editor will be specifications, paper prototypes, and individual computer prototypes of several primitive instructional support tools.

Year 2: Design and Development

The Peer. During year two we will release the Peer's source code and documentation for free download according to the Open Source software model. We will continue stabilization and development work in harmony with user studies of the software and user-contributed patches and enhancements under the Open Source software model. Additional functionality allowing the Peer to communicate with the Recommender (e.g., the ability to send usage statistics and receive recommendations) will be added during year two. The **result of the second year's work** on the Peer will be a more stable and usable Peer that can pass automated and adaptive recommendation of relevant resources on to users.

The Recommender. During year two we will implement the selected collaborative filtering algorithm, implement an appropriate database design, complete the necessary data collection services software (e.g., the ability to harvest usage and optional user profile information from the Peer), and provide a method of recommendation transport from the data center to the Peer. This source code and documentation will also be released according to the Open Source model. The **result of the second year's work** on the Recommender will be a functioning recommender system integrated with the Peer.

The Editor. During year two the individual prototypes of the Editor's tools will be integrated into a standalone application capable of utilizing arbitrary resources on the user's personal computer. An open specification for the creation of additional instructional support tools will be published, allowing others to contribute new tools to the system. Source code and documentation for the freestanding Editor will be published according to the Open Source model. The **result of the second year's work** on the Editor will be a fully functioning, standalone application facilitating the instructional use of educational resources on users' personal computers.

<u>Year 3: Design and Development</u>

The Peer + Recommender + Editor = EduCommons. During year three the final integration, evaluation, revision, and testing of the fully integrated EduCommons tool will occur. The standalone Editor will be integrated into the Peer in order to provide "one-stop shopping" for educational resource sharing, discovery, and use. Final versions of clean, documented source code and other documentation will be prepared and released under the Open Source model. The **result of the final year's work** will be a free, open source, peer-to-peer system for sharing, finding, and using educational resources to help improve the quality of teaching, learning, and research.

<u>Years 1 through 3: Evaluation</u>

Project evaluation will include formal and informal evaluation methods that mirror the three development phases of analysis, design, and implementation. Formal evaluation methods and subsequent reports will include: a needs assessment to identify intended users and their needs and issues with respect to the sharing, discovery, and use of educational resources, including incentive structure and intellectual property studies; a descriptive case study detailing qualities of project design, implementation, and analysis; expert reviews; and a comparative analysis of post-implementation use and perceived utility of the system. Informal evaluation methods consistent with developmental evaluation design will include participant observation of focus group design meetings, and interviews with resident experts and end-users.

External reviews by specialists in the Department of Instructional Technology at Utah State University will evaluate prototypes for quality of system interface, interaction (location and combination of learning objects), and support. The project team will assess system utility qualitatively through case studies and quantitatively through focused e-survey questions to system users. Results from these descriptive observations, surveys, and reviews will continually inform developers during project revision and implementation cycles.

DISSEMINATION AND SUSTAINABILITY

Results from the research will be disseminated by the investigators in professional journals (for example, *Educational Researcher, ACM Transactions in Computer-Human Interaction (ToCHI),* and the *Journal of Interactive Learning Research*), conferences (for example, the annual meetings of the American Educational Research Association and Association for Educational Communications and Technology). Additionally, project source code along with accompanying documentation and white papers will continue to be freely available under licenses like the GPL and OPL in the spirit of the open source movement.

Additionally, several sustainability mechanisms will be explored. As the collection of recommendation and usage metadata grows within the EduCommons, these data become a valuable resource in their own right. The Recommender / data center database contains data about which resources users perceive to be of value. As such, we will explore licensed access to our Recommender database by other educational digital libraries and commercial vendors. We will also explore the possible commercialization of the EduCommons system itself.

RESULTS FROM PRIOR NSF WORK

We are currently engaged in work on NSF award DUE-0085855, “The Instructional Architect: A System for Discovering, Recommending, and Combining Learning Objects.” This award was made in the amount of $399,788 for a period of two years, beginning in September of 2000. The project involves the creation of a service layer for existing digital libraries that allows resources within the libraries to be more effectively discovered and utilized. The project was spurred by interactions with NSF-funded digital libraries for educators. These libraries provide access to high-quality resources and a simple search feature, but stop there. Because these systems do not provide users with the ability to combine and utilize resources within the system, users are currently forced to download resources offline, combine them by hand in an HTML editor, and republish the new collection on the web. This degree of necessary technical expertise keeps many educators from effectively utilizing the resources within the libraries.

Accordingly, we proposed the creation of (1) an online learning object combination environment, and (2) an advanced search / collaborative filtering feature that would both recommend to users additional resources and like-minded people. These mechanisms relate closely to those in the current proposal, and will inform this work significantly.

Although project funding just began in September 2000, a prototype of the Instructional Architect system is already available for download, complete with source code but missing some functionality (such as the recommender system; see http://ia.usu.edu/). While the study is not yet complete, a publication describing the Instructional Architect system is available (Recker, Walker, & Wiley, 2000), as are white papers describing various aspects of the project (e.g., describing user-selected metadata elements for educational resource discovery). See the project website, http://ia.usu.edu/.

While the Instructional Architect study is similar to the proposed EduCommons study significant differences between these proposals exist and should be pointed out here:

1. The digital libraries to be serviced by the Instructional Architect project are web-based. This is significant because the infrastructure for accessing the resources already exists, largely without design considerations for the learning domain. Additionally, while the Web was designed to be easy to use, publication of resources on the web remains difficult, often involving Unix accounts, mastery of the file transfer protocol, the differences between ASCII and binary file formats, et cetera. The EduCommons study will create a new infrastructure designed specifically with learning as its primary goal, with an appreciably lower barrier to publication and participation.

2. The digital libraries to be serviced by the Instructional Architect project are curated collections, tightly controlled for quality. While this approach has obvious benefits it also serves to propagate resources approved by and useful to the majority of users, often underserving or underrepresenting certain populations. The EduCommons proposal addresses these issues by (1) allowing members of all populations to publish resources (thus increasing the diversity of available resources) and (2) providing advanced resource discovery mechanisms, such as the Recommender, which will allow people to find satisficing resources, even when the collection is large and unmoderated.

3. Closely related to point two, the digital libraries to be serviced by the Instructional Architect depend on collections development for their resources. In other words, these libraries cannot grow unless (1) someone creates new resources for the library, and (2) curators inspect and approve the resources. Thus the process of maintaining the library with the most up-to-date educational resources is costly and time consuming. The EduCommons system harnesses the educational material which already exists on the personal computers of teachers, students, and researchers. Each year that new students take new classes, or that researchers do new research, there will be new material in the system. The EduCommons system leverages existing resources by making them accessible, discoverable, and instructionally reusable.

Other information relating to the project is available online at http://ia.usu.edu/

CONCLUSION

The proposed system harnesses the instructional power of existing learning and research resources by fusing two emerging technologies: peer-to-peer and learning objects. We believe that the proposed tools could radically advance learning and research by providing users with ubiquitous access to high quality content and high quality tools for manipulating the content. We also believe that our broad partnerships and open source approach will afford the greatest project impact possible.