Planning Grant for the Use of Digital Libraries in Undergraduate Learning in Science

Kurt Maly Mohammed Zubair Stewart Shen Steven Zeil

Old Dominion University
Dept. of Computer Science

A. Project Summary

In 1999 we intend to submit a full proposal in which we will propose to develop a comprehensive, easy-to-use digital library and a process for undergraduate science and engineering education. In preparation for this proposal, we will investigate the following questions in this planning grant: What features are required in a digital library to support active learning? What types of tools can be used with a digital library to enhance the learning experience for undergraduate students? Does the use of such a digital library increase learning in undergraduate science and education? How can we improve the dissemination and reuse of successful courses?

Many projects have been successfully completed in the arena of delivering learning in Higher Education over the Web and the field of creating digital libraries. Web technology is used to deliver course material to students and to support collaboration among students, and between students and the teacher. However, it is not clear how such an investment can be preserved and enhanced for future use.

The purpose of the future, full proposal is to develop and make widely available a student-centered, active-learning mechanism in undergraduate science and engineering curricula which relies on digital libraries and simple, web browser-based, essential collaboration tools. Digital libraries will be used in two major ways. On the one hand, students will learn research techniques, information reuse, and documentation and reporting practices not generally introduced until graduate school. On the other hand, we intend to demonstrate that we can publish the essence of a student-centered course, be it delivered over the Web or through more traditional means, into a digital library, together with material associated with a particular course instance. This will allow for subsequent retrieving, reviewing, and adapting of the material for new instances of the course in potentially different environments. Two events have moved us to propose to address these issues. The first is the TechEd initiative in the Computer Science department at ODU and the second is the NCSTRL+ project between ODU and NASA Langley. The TechEd initiative will offer an entire undergraduate program using technology-based distance learning methods to realize an active-learning paradigm. NASA is funding the development of NCSTRL+ which will provide the ability to classify material and associate appropriate terms and conditions with parts of the material such as who can access what. We expect to have access to a prototype which has these capabilities and a reasonable corpus to allow for the meaningful research expected of undergraduates in selected CS courses.

In order to write such a proposal, a number of questions pertaining to the design of such an environment need to be investigated. In particular, what effort level is required to publish into and maintain such a library? Are using digital libraries for research techniques an effective means of increasing the rate of or improving learning in science and engineering courses? Is a digital library a cost-effective way to store and organize teacher's and students' material? What should be stored? What form of a digital library would be best? Hence, for the planning grant we propose to develop prototype tools, processes, and an environment by analyzing three active-learning delivered CS courses; to instrument the prototype tools and the digital library, and run one CS course and collect data. We shall analyze the results and write a full proposal for a system and a process to support an active-learning paradigm with the help of digital libraries, the key benefit being the availability of the course for future teachers.

C. Project Description

Instructional methods in academe are shifting from a teacher-centered paradigm to a student-centered paradigm [8]. In this new paradigm, the student becomes an active participant in the class, and peer collaboration becomes an important component in the learning process. Advances in networking, digital libraries, and digital media technology together with the growth of the Internet are making the web an effective framework for supporting this type of active learning.

Very successful projects funded by NSF and DOE have demonstrated various innovative approaches to instruction. Fox [13], for example, has created a number of on-line courses where students are the primary focus, and the teacher's main role is to organize material and make it available on the web. Technology is thus used to enable the students to review multi-media material and to create such material for the benefit of the class. Technology is also used to support collaboration among students, and between students and the teacher. However, most of the publishing is done to the Web rather than into a digital library. This results in all the problems the Web community is only too familiar with: 'link rot and content rot' as well as the difficulty of precise and complete location of resources. A considerable effort is under way in the Web community, in particular, W3C, to improve content labeling, indexing and searching [2]. However, very little effort is being made to coordinate these efforts with the digital library community, for instance the DLI initiative supported by NSF, where many of the problems of creating and maintaining specialized collections have been addressed.

After so many successful projects of this nature, why are students at the "average" college not sharing in these improved learning experiences? Why are digital library publishing not used in the classroom more often as opposed to web publishing? From anecdotal evidence, we can state that typically after completion of successful projects, only project faculty continue to use the newly-developed approaches and materials, and no other faculty members at the local institution let alone at other colleges benefit from those results. This failing is due largely to: 1) lack of easy-to-use guidelines for using digital libraries, 2) insufficient support for multi-disciplinary, multi-formatted content, 3) insufficient capabilities in existing, suitable digital libraries to support *terms & conditions*, which are essential for publishing confidential course material, and 4) a shortage of simple and intuitive tools for both teachers and students to support the process of publishing, finding and using material in digital libraries appropriate for undergraduate science and engineering courses.

In the full proposal we plan to write as part of this planning grant, we shall look at how digital libraries can help to increase the effectiveness of teaching and learning in undergraduate science and engineering curricula at the average college, where instructors are faced with heavy teaching load, students frequently have work commitments, and where only a limited infrastructure provides access to Internet resources. In particular, we will look to the ways that digital libraries can support:

Archiving and retrieval of course materials We will seek to provide an environment and a process for preserving courses and their instantiations, thus making courses available for later teachers. This environments will be initially populated with a variety of excellent existing tools, methods, and other informational content available both to teachers to create interesting courses and to students who participate in such courses. In particular, we will 1) develop software tools to support collaborative use of research methods, 2) organize informational content in various media and formats from CS research, NASA Langley research, graduate theses from CS, and from NSF CISE educational projects into one national digital library,

Instantiating course offerings from archived material We will seek to develop usable processes for collecting archived material from a variety of sources in the library to form a particular course offering, together with material specific to the particular instantiation (offering) of the course.

Presentation of course offerings We will explore the role to be played by digital libraries during the actual teaching/learning process, with particular emphasis on supporting student research under an active learning paradigm and on the use of collaborative exploration tools in teacher-student interactions and in student group activities.

We will propose to demonstrate the utility of digital libraries in these 3 roles with actual courses from the TechEd initiative at ODU, and will attempt to assess the success of delivery of this type of courses in engineering and computer science.

Just to write such a proposal, a number of questions pertaining to the design of such an environment need to be investigated. In particular, 1) Is a digital library an effective way to store and organize teacher's and students' material, and what is to be stored? 2) What effort level is required to publish into and to maintain such a library? Are students and teachers likely to find this time- and cost-effective? 3) How difficult is it to be used by faculty and students with varying technical backgrounds? 4) What organizational structure in a digital library will improve user satisfaction? 5) How can we measure the effects of student research via digital libraries on the rate of or improved learning in science and engineering courses? 6) Will students perceive an advantage in publishing their personal course portfolio into a digital library's persistent canonical format, as opposed to creating a personal website?

We propose in this planning grant to develop a set of prototype tools and processes, to elicit requirements for the complete library system, and to do a proof of concept study to answer, at least preliminarily, these questions. The key concept we intend to demonstrate is that we can publish the essence of a student-centered course, be it delivered over the Web or through more traditional means, into a digital library together with material associated with a particular instance (course offering). Crucial to this effort will be a digital library that provides for the ability to 1) publish a variety of course-related materials, 2) allow subsequent retrieving, reviewing, and adapting of the material for new instances of the course in potentially different environments, and 3) classify material and associate appropriate terms and conditions for accessing and updating different parts of the material. NASA is currently funding a project in the Computer Science department at ODU (the PI is the same as for this planning study) to develop NCSTRL+ for use in NASA's STI effort. We expect to have access to a prototype which has these capabilities and a reasonable corpus to allow for meaningful research expected of undergraduates in selected CS courses.

Hence, for the planning grant we propose to develop prototype tools, processes, and an environment by analyzing three active-learning delivered CS courses; to instrument the prototype tools and the digital library; and to run one CS course while collecting data. We shall analyze the results and write a full proposal for a system and process to support an active learning paradigm with the help of digital libraries, the key element being the availability of the course for future teachers.

1 Background

1.1 Using Digital Libraries

Digital libraries (DLs) store materials in an electronic format, ensuring certain quality standards, and making information accessible for searching and retrieval [34]. They are usually implemented at a level above the web, utilizing it as their infrastructure and transport. Traditional WWW search engines (Yahoo, Alta Vista, and the like) are promoted in terms of the completeness of their coverage of the web, and so are biased heavily towards the high recall end of the recall/precision tradeoff. The ever-growing noise levels in their search results make them increasingly less suited to targeted research and education needs. In contrast, users of digital libraries over the web can systematically search stored materials according to various criteria and obtain results with high precision depending on the quality of the DL searched. The U.S. Government created the Digital Libraries Initiative [35], the flagship research effort for the National Information Infrastructure (NII), which seeks to bring highways of knowledge to every American [22].

There are a number of significant digital library projects, aimed at K-12 education through university research. These include EduPort [23], with a centralized DL stored at the University of Nebraska and used for K-12 students, a DL of "algorithm animations, video, images, and text" maintained by Virginia Tech and NSU [14], the National Digital Library Program [37], the Image Engine HPCC Project [20], the Virtual Hospital [38], and the Digital Library System for Periodical Distribution [3], and NCSTRL [7].

NCSTRL is one of the first successful, operational, distributed DLs on a national scale. It is a library of computer science technical reports which covers a significant portion of the current CS research. Over one hundred publishing organizations in three continents [26] currently make up this collection, and this number is continuously growing. This library consists of a body of about 10,000 reports, high quality standards, and a management process for publishing into the library. The archives of the reports are maintained at the publisher cites with bibliographic information in support of searches being maintained at regional servers. Mirror sites of the index information are maintained across continents. The Dienst digital library protocol [5] is the engine used to support both NCSTRL and our related project, NCSTRL+ [29].

There are commercial digital library offerings with support for rights management, such as the IBM Digital Library [16], but such such products are architecturally closed. A community standard for DL interoperability has yet to emerge, so we have chosen to avoid potential proprietary pitfalls.

1.2 Roles for Digital Libraries in Education

We envision in this project, three primary roles for digital libraries in the educational process: as an archive for course-related information and prior course materials, as a tool for constructing and organizing materials supporting a specific offering of a course, and as a tool in an active (possibly virtual) class session.

1.2.1 Archiving Course-Related Information

A large investment is being made at different universities to offer Web based courses for asynchronous distance learning. There is a need for digital libraries that can protect this investment. We need to archive various instances of these Web based courses with all the students related material (projects, assignments, examinations, etc.).

1.2.2 Instantiating Course Offerings

Although a digital library may aid the instructor by providing access to substantial amounts of potential course material, the instructor still faces a daunting task in organizing the selected materials in support of a specific course. In addition, a framework must be established for accumulating offering-specific information (student submissions, grades, student questions & e-mail, etc.). Even a single course may, furthermore, feature multiple concurrent offerings, possibly by different instructors. In a partially asynchronous learning environment (such as the ODU TechEd initiative), course offerings that started on different dates may continue to exist simultaneously.

The materials for a given course offering are likely to span a wide variety of media and data formats. These will be interrelated, often at a semantic level inaccessible to automated analysis. As courses and curricula evolve, the challenge is to maintain multiple distinct consistent configurations of course materials. Most importantly, the processes for doing so must be usable by instructors with widely varying technical backgrounds and level of motivation.

1.2.3 Supporting Active Learning

Active learning as a concept is based on the constructivist view of teaching and learning, which holds that all learners must engage in an active role in constructing their own knowledge and that they are not merely empty vessels to be filled by sage professors. This view of learning acknowledges the importance of the role students play in the teaching-learning process. Students need to engage in the creation of their own meaningful view of the discipline, and professors need to value their views and their participation in exploring that discipline. Students should be provided structured guidance in the exploration of a topic and increasingly left to their own devices to develop the mental constructs essential to understanding that discipline. The student is considered a vital participant in an academic environment where the student is seen as a thinker, a creator, and a constructor [15]. The primary vehicle for learning is considered research through resource discovery. [30] The professor becomes more of a facilitator rather than an instructor and students become co-learners. Competency in specified objectives is obtained through simulations and activities and therefore the best test is one which checks whether a learner is using what she has learned in a particular course.

Class sessions under an active learning paradigm are likely to place heavy emphasis on collaborative effort. Collaborative access to digital libraries, however, cannot be readily achieved with standard browsers. What one person does with such a browser is shown on his or her computer screen only and can be seen only by those physically present. In order for instructors and students to work together at different locations, group collaboration tools are needed. Recently, group collaboration research has produced sophisticated environments allowing researchers to collaborate over the Internet. A number of such systems are available both commercially [32, 31, 25] and as research prototypes [21, 10, 36] from national laboratories and universities.

1.3 Limitations of Current Systems

1.3.1 Protecting Intellectual Property:

The development of on-line course materials represents a significant investment. Protection of this investment is often a practical prerequisite for digital publication of the material. One of the issues here is the *terms and conditions* of the content, that is, what material can be accessed by whom. Most of the digital library efforts are not geared to handle this kind of support for the content. This support has to be made available at all levels of a digital library, namely: publishing, managing, and searching & retrieving. Such an archive will be beneficial to the faculty responsible for future offering of the course, and to the students.

1.3.2 Web versus Digital Library:

Why do we consider publishing on the Web insufficient for a learning environment and believe that digital libraries are essential for supporting an active learning paradigm? Digital libraries can be considered islands of specialized collections on the Web which have their own management policies for publishing and access control. Being libraries they take on the responsibility for preserving material solving the "link rot" problem so commonly found in the Web. Secondly, digital libraries provide (and enforce) the use of metadata which describe the content of published material and allow for accurate searches, again a major problem in the current Web search engines. The current strong efforts in W3C to define standards for XML [2] and RDF [19] will provide a boon for making different digital libraries interoperate and will allow them to export their metadata format to other libraries. For example, an instructor can publish material specific to a course into a library, including links to other libraries whose metadata is made accessible to student searches. The instructor may also allow students to submit their work (which may involve a variety (for example an anthology of many different small pieces of content: letters, video clips, simulation programs) for a course by publishing it in the library, where the instructor manages and controls subsequent access to these works. Students might be prohibited from viewing one another's submissions, but retain access rights to their own work for the purpose of building a personal portfolio of their entire University work.

1.3.3 Undergraduate Research Methods:

Current DLs are almost uniformly single-domain. However, in undergraduate science and engineering curriculum a great deal of overlap exists between subjects studied in a variety of scientific and engineering disciplines. The multi-disciplinary nature of today's research and industrial projects require participants to be conversant in disciplines other than their own. Secondly, for operational economy of scale there must be a mechanism for combining small discipline-centric archives together into larger archives, such as combining all the DLs of various university engineering departments into a DL for the entire college of engineering, or other appropriate organizational entity. This should be accomplished without impacting users who need to restrict their searches to a specific discipline. In addition to this limitation, frequently a gap exists between the publisher and the consumer of informational content. The producer of informational content has a wealth of material in many different formats and with different aims of consumption. For instance, a document needs to be read but source code needs to be installed, compiled and run. Current DLs, including NCSTRL, do not support the publication and interpretation of such diverse information objects.

1.3.4 Configuration Consistency

Because current digital libraries tend to specialize in a limited number of predetermined information formats, the configuration and version management issues outlined in the previous section tend to be considerably less severe. Indeed, the configuration problems to be faced in supporting course archiving and instantiation would seem to be more reminiscent of those faced in repositories for Software Engineering Environments (SEE's) than than in prior digital libraries, which have also had to deal with managing substantial networks of diverse, interrelated, and evolving objects.

Configuration dependencies in SEE repositories have typically been captured using elaborate, programming-language-like, type systems to describe stored objects and the "type signatures" of tools that are available for operating upon them. Experience had shown that this approach is powerful enough to facilitate the publishing of both new and legacy tools and data [9, 40]. On the other hand, this approach's analogies to programming practices may prove inappropriate for non-computer scientists publishing into a digital library.

1.3.5 Collaboration Tools:

Sophisticated collaboration environments require either broad bandwidth, expensive equipment, expensive software, free but unsupported software, or do not scale to many users over a congested Internet, a lethal set of choices for the students and teachers of our target audience. Even existing, simple systems have specific problems. For example, Yarn [33] operates only in conjunction with the Mosaic browser, which is now out of favor and is somewhat inadequate for today's WWW. On the other hand, the EMSL WebTour [10] requires private (non-WWW) socket connections as well as connecting to the API's of Netscape and/or Mosaic browsers. The major limitation with these systems is that they require additional software along with a standard browser. We need an approach that does not require any additional client software or any modification to the browser.

2 Future Full Proposal - Outline

The key problem we wish to address is how to enable faculty in colleges with limited resources to offer courses in which students, also with limited resources, can learn through active participation and peer collaboration. We shall look at how digital libraries can help to increase the effectiveness of teaching and learning in undergraduate science and engineering curricula at the average college, where instructors are faced with heavy teaching load, students frequently have work commitments, and where only a limited infrastructure provides access to Internet resources. We will examine the use of digital libraries in the three principal roles described earlier: *archiving*, *instantiating*, and *presenting* course material.

To support these roles, we will propose to develop a well-populated digital library, collaborative access tools, and processes/templates for working with the library in an educational setting. The design of these components will be driven by the search for solution to the following problems:

- Traditional libraries are insufficient in terms of the kind of material available for undergraduate research, in cost of delivery and in the time it takes for delivery;
- Existing digital libraries are usually single-domain, single-genre, single-type-of-organization. In science and engineering, many disciplines overlap with others; for example, an airflow problem may have relevant material in both a NASA and computer science DL;
- Existing digital libraries do not support terms and conditions for controlling access to material which is vital in a education environment where confidentiality of student records is important;
- Few digital libraries in academe have critical mass involving enough publishers or users;
- Collaborative tools need to work in congested networks and scale well. In addition, they should pose minimal requirements on the network support infrastructure;
- Having students use research in undergraduate courses is presently an expensive proposition, expensive in instructor efforts and wasted students' time;
- The latest research results are not easily accessible unless one knows who has conducted the research, where it is located, and in what form it exists. The student typically does not know any of this information; and
- The faculty lack access to templates or processes that can be used to build new courses based on existing archived courses in the digital library.

2.1 Objectives

The purpose of this proposed work is to provide a student-centered, active learning mechanism in undergraduate science and engineering curricula by developing a digital library environment supporting course archiving, instantiation, and presentation. To accomplish this purpose, we will propose to:

- Develop software tools based on NCSTRL+ to use and publish in a multi-domain, multi-genre, multi-organization, terms & condition digital library targeted for undergraduate science and engineering curricula.
- Develop software tools, including a coordinated browsing system, for support of collaborative use of research methods.
- Publish informational content in various media and formats from CS research, NASA Langley research, and from NSF CISE and DOE educational projects into searchable entities in NCSTRL+.

- Develop generic process descriptions of methods for publishing and maintaining course material (both teachers' and students') into a digital library, as well as material connected with actual instantiations (offerings) of these courses.

- Demonstrate this approach with courses, delivered in a student-centered learning paradigm, from the TechEd initiative at ODU.

- Assess the effectiveness of these approaches in courses at ODU in the classroom, office and students' homes with the tool environment and DLs we have created.

- Disseminate the results.

The detailed steps to be taken in achieving these objectives will depend, of course, upon the answers obtained from the planning grant project.

3 Planning Grant, Proposed Work

In preparation for the future full proposal outlined in Section 2, we propose, for this planning grant, to investigate the questions: 1) Is a digital library an effective way to store and organize teacher's and students' material, and what is to be stored? 2) What effort level is required to publish into and to maintain such a library? Are students and teachers likely to find this time- and cost-effective? 3) How difficult is it to use by faculty and students with varying technical backgrounds? 4) What organizational structure in a digital library will improve user satisfaction? 5) How can we measure the effects of student research via digital libraries on the rate of or improved learning in science and engineering courses? 6) Will students perceive an advantage in publishing their personal course portfolio into a digital library's persistent canonical format, as opposed to creating a personal website?

To provide preliminary answers and allow us fully to understand the issues, we propose:

- To perform a requirements analysis with 3 active-learning CS courses and

- To develop a subsequent prototype framework for teaching and learning with a digital library. This framework will include collaboration tools, NCSTRL+, process tools and a supporting environment.

- To instrument the prototype tools and the digital library and to run one CS course to collect data.

- To analyze the results and to write a full proposal for a system and process supporting an active learning paradigm with the help of digital libraries and preserving courses for future offerings.

3.1 Summary of NCSTRL+

Old Dominion University and NASA Langley Research Center are developing NCSTRL+, a digital library with particular emphasis on the storage of multi-discipline and multi-genre information. NCSTRL+ is based on the Networked Computer Science Technical Report Library

(NCSTRL) [5], which is a highly successful digital library offering access to over 100 university departments and laboratories since 1994, and is implemented using the Dienst protocol [6]. The NASA/ODU team is implementing NCSTRL+ using Dienst instead of other digital library protocols such as TRSkit [27] because of Dienst's success in several years of production in NCSTRL. Dienst appears to be the most scalable, flexible, and extensible of digital library systems we surveyed [11]. Dienst also serves as the basis for other focused digital library projects, including: the Electronic Thesis and Dissertation Project [12], the University of Virginia undergraduate engineering thesis project [39] and the ACM SIGIR conference proceedings project [1].

NCSTRL+ makes two main contributions: clustering and buckets. Buckets [28] are an object-oriented construct for creating and managing collections of logically related information units as a single object. A bucket can contain both different data syntax, e.g., PostScript, PDF, Word, and different data semantics, e.g., manuscripts, data files, images, software. Buckets are similar in concept to the "digital objects" first proposed in [17] and later developed in the Warwick Framework [18] and FEDORA [4]. Buckets provide two main functionalities: aggregation and intelligence. Aggregation is achieved by allowing the user to group logically related data items together in a single container construct and to treat the collection of objects as a single entity. Buckets provide the mechanism, not the policy, for expressing relationships between objects. Buckets also act as intelligent agents by allowing communication between other buckets, people (owners, users, etc.) and arbitrary network resources. Buckets communicate via a blackboard-like system known as the the "Bucket Matching System". Buckets, not archives or repositories, are responsible for negotiating presentation and terms and conditions with users. It is the inclusion of archive-type and intelligent agent functions with the object itself that differentiate buckets from the Kahn-Wilensky digital objects and their derivatives.

Clustering provides a way to partition collections along various logical dimensions. NCSTRL+ defines 4 clusters: subject category, publishing organization, archival type, and terms and conditions. Subject categories allow multi-discipline DLs to be built, by allowing users to select which disciplines they would like their searches to include, for example, aeronautics, space science, mathematics, computer science, or physics. NCSTRL+ has selected the NASA indexing method of subject matter as opposed to the too unwieldy (for the average viewer) Library of Congress subject classification system. NCSTRL+ also allows for inclusion of other classification schemes, such as the ACM subject classification scheme, as a subdirectory in the NASA classification. The publishing organization identifies the organizational entity responsible for publishing, whether it be a laboratory, department, or a professional organization. The archival type cluster defines what kind of information is included: project reports, technical reports, software, datasets, journal articles, courseware, theses, patents, personal. Associated with these types will be a description which gives an indication on how these objects will be archived: how long they will be preserved, what presentation tools are available in the archive and what machines (real or emulated) can run these tools.

A sub-category of this cluster specification will allow the specification of the time the material has been entered and the frequency of its access. Implicit in this classification is the complexity of the publication management process, that is, presumably the longer a bucket is to be preserved the more rigorous the approval process for publication. This cluster is especially important when the DL grows to include material beyond traditional publications, such as course notes, course software, and materials generated collaboratively. The terms and conditions cluster allows users (both publishers and viewers) to select which classes of access control and monetary requirements the search should include. This cluster impacts beyond the traditional "who can read what" restrictions. Since buckets are dynamic objects, there have to be mechanisms for allowing restrictions to be placed on who can update and maintain them. Some buckets may reveal part of their contents to certain user classes but not all their contents. Terms and conditions will allow specification of such items as with whom the bucket can communicate and from whom it can accept content and architectural updates. The major categories NCSTRL+ will support are: Classified - requires SSL and special code; Groups - requires membership in a group with password; Cost - varying subcategories ranging from flat fees to per-access charges; Copyright - restricted to proper use as per copyright notice; and Unrestricted.

3.2 A Digital Library Environment for Undergraduate Courses

A canonical digital library capable of both publishing and retrieving buckets of multimedia research and teaching material in multiple disciplines is in itself a compelling goal. When coupled with an environment of tools and procedures supporting archiving, instantiation, and presentation of course material, it promises to be a valuable national resource for undergraduate technical education. Faculty will be able to prepare for classes with more ease, higher quality materials can be easily included into curricula, and student-centered learning can be more readily incorporated into courses. To facilitate the future construction of such an environment, our major proposed activities involve requirements analysis, development of software prototypes, and evaluation and analysis. In this section, we describe further the requirements analysis and prototyping activities. Evaluation and analysis is described in Section 3.3.

3.2.1 Archiving Web Based Courses:

A large investment is being made at different universities to offer Web based courses as part of asynchronous distance learning initiatives. It is not clear how a Web based course that has been offered will be archived along with all the students' related material such as projects, assignments, and examinations. The archiving will be beneficial to the faculty responsible for future offering of the course by having a template of the course. For students, it will give them not only a perspective as to the material covered in the course but what actually happened in previous offerings. Administrators will be able to evaluate faculty and be responsive to accreditating agencies. For all these people involved with the archive one important concern is the issue of terms and conditions, particularly, as it relates to confidential student material such as grades.

The Computer Science department is currently finalizing detailed plans for offering an entire undergraduate program using technology-based delivery methods to realize an active learning paradigm. The program, TechEd, will support a spectrum of teacher-student interactions ranging from mostly self-paced to mostly interactive in group sessions. These will be supported by learning technologies ranging from IRI [21], an advanced, real-time, multimedia, collaborative environment, to more conventional Web presentations enhanced with video conferencing capabilities [24]. The TechEd initiative will be an ideal testbed to explore our ideas of using digital libraries. We propose to use parts of ODU's funding for this initiative as a matching grant to both the planning grant and a future full proposal.

We will focus on studying the archiving issues related to three courses that are part of the TechEd program. The three courses have been tentatively identified as: Computers and Society (CS 300), Data Structures and Algorithms (CS 361), and Digital Libraries (CS 496/596). The first of these is a non-programming course heavily attended by non-CS majors. The second has a significant individual programming component, and the third places heavy emphasis on group project activities. The objective in selecting these courses was to cover as many issues as possible in a short feasibility study.

A list of issues to be resolved during requirements analysis includes:

Bucket granularity Buckets are recursive digital objects and the organization of buckets may depend on the viewer, i.e., an instructor may want to have a different organization of the material. Relevant NCSTRL+ features: bucket aggregation tools, publication management process tools

Completeness of buckets Buckets will contain both static course material and dynamic material relating to course activities such as running a program and analyzing its complexity; the issue is how self-contained a bucket has to be. Relevant NCSTRL+ features: environment description in which buckets and their packages can be viewed.

Student portfolios Students may want to create portfolios of their work in courses they have taken for reference and reuse in later courses or for future job searches Relevant NCSTRL+ features: aggregation tools, archival type, subject classification.

Course administration What is the process for supporting indigenous methods for administering courses and evaluating them. The issue is not to provide a generic method but rather to allow for publication of the results of these methods for later retrieval; for instance, a faculty member wants to provide supporting evidence for a promotion case. Relevant NCSTRL+ features: multi-format support for bucket creation.

Role dependent access Different people with different level of expertise will want to access course information and research material; user interface should be oriented not towards functionality but be situation oriented: for research purpose, reuse of course, preview of course to be taken by student, evaluation, accreditation. Should each user be given a 'library card' which details preferences and capabilities of users? Should the interface provide alternatives depending on system capabilities of user, preferences of user, and content user is trying to access? Should the interface provide collapsed hierarchies, overviews, previews, integrated site maps,scenarios, examples? This is a major research issue for NCSTRL+ development and will only be addressed in an exploratory way.

Course evolution Courses may evolve in response to curricular pressures, leading to a variety of related, but incompatible course offerings that may be concurrently populated by different groups of students, who will be largely unaware of the existence of these alternative versions. How are these varied configurations tracked, identified and checked for consistency? What processes support identifying not only the appropriate version of a single course offering to present to a given student, but the appropriate series of curricularly consistent offerings as that student progresses through a degree program? Is the NCSTRL+ bucket/clustering model sufficient to this task or is a higher level model required? Relevant NCSTRL+ features: buckets & clusters.

3.2.2 Instantiating Web Based Courses:

As courses are actually offered, a new set of issues arises that is less related to the mechanism of archiving than to the ease with which archived information can be identified, rearranged, and repackaged.

As discussed in Section 1.3.4, these include issues of maintaining multiple consistent views of courses while still supporting course maintenance and curricular evolution.

Requirements analysis for this digital library role will probably be mainly concerned with developing processes and templates for use in educational environments. Also of interest is the suitability of the NCSTRL+ bucket mechanism in support of these requirements, or, alternatively, the use of that mechanism to implement more abstract models if such are dictated by the discovered processes.

A preliminary list of issues to be addressed under this role includes:

Course review and reporting What is an appropriate process for creating sample materials from actual courses (e.g. accreditation requirements: sample of x students' exams, homework, ...)? Significant issues include privacy and obtaining releases from students and faculty to publish records of actual events during a course offering. Relevant NCSTRL+ features: terms & conditions.

Course maintenance How can an instructor readily update existing course materials? How does an instructor indicate what materials from existing courses (possibly prepared by other instructors) should be reused in a given course. Relevant NCSTRL+ features: bucket matching and intelligence.

Role dependent access (See above)

Course instantiation What processes support the instantiation of a course as a particular offering? This includes both the publishing of selected course materials to define this instance and the granting of access and other privileges to a moderately large group of previously unknown users (the students registered for the offering).

Component publishing Different instructors have different methods for handling assignments, exams, quizzes, projects, and group discussions; what basic processes and features are needed to support them? Relevant NCSTRL+ features: terms & conditions, archival type; relevant tools: collaboration tools.

3.2.3 Presentation of Digital Library Material:

The traditional model of individual access to digital library contents is too limited in comparison to the activities that may arise when these libraries are employed as part of an active learning curriculum. Some possible scenarios in a teaching environment are:

Groups of students: Courses that are project oriented may require a group of students to collectively research or explore information on the Web. For example, a group of students may be working together to write a report on the subject of computer security. For this, students may decide to research the material together to identify papers which they can use for writing the report.

Groups of students and an instructor: When an instructor holds a recitation session with a group of students to explain or elaborate on some concepts discussed earlier in the class, an instructor may find it useful to use different media such as animation and video available in a digital library to explain a breadth-first search algorithm.

An instructor and a student: When one student wishes to interact with the instructor with some specific questions, the instructor may find it appropriate to co-browse material from the lecture for that course on the Web with the student.

Collaborative web browsing allows a group of users to surf the web together. The users could be geographically distributed and possibly working on different platforms. In this environment when a user in the group loads a new document from a site; the same documents gets loaded on all the other users' web browsers. One possible approach is to use technologies based on the X Window System protocol. However, this approach is not amenable to a group of users working on multiple platforms like Unix, NT, MacOS, and Windows95. We have designed a collaborative browser that allows co-browsing on any platform with a standard graphical web browser (for example, Netscape Navigator and Microsoft IE) without requiring the installation of specialized plug-ins or other extensions.

During the planning grant period, we will implement the basic collaborative browser capability. The full system (at the end of the future project) will have additional features such as asynchronous support. An instructor will be able to create a scripted tour of several Internet sites and register this tour with the proxy server. Students can then access the series of WWW locations in the path determined by the instructor. We intend to support either text or audio annotations for the pages placed in a tour. An asynchronous collaborative ability will facilitate the creation of courseware materials that can be reused. It also allows greater flexibility for large groups of students working together whose respective time constraints (work, parenting, etc.) would otherwise prohibit traditional collaboration methods. We also intend to port the "hands-free" audio facility[1] of IRI [21] to work with any browser. This feature will enable C-browsing participants to have an audio discussion at the same time they are browsing the web.

The collaborative browser is targeted for development within this early planning phase because of its utility in the CS course to be run using the prototype library. The list of open requirements questions is consequently somewhat shorter, but includes:

Student-Instructor interaction : A variety of interactive modes will be required to support a complete curriculum. How effective are available tools, including the co-browser, in supporting instructors' varying styles of teaching? Relevant tools: collaboration tools.

3.3 Evaluation and Analysis

In the preceding discussions, we have raised a number of issues regarding the potential impact of digital libraries on education in science and technology. We can summarize these in what we shall call the *DL premise*:

> Digital libraries are useful in enabling the use of technology-based delivery methods for courses and in enhancing active learning.

[1] Using silence detection, audio mixing, and distribution to a group.

During the planning grant period, we will be offering one CS course from among those described in Section 3.2.1. This course will make heavy use of the prototype library, and will be the arena for the collection of metrics related to this hypothesis.

In general, our evaluation approach for both (planning and full) projects will consist of transaction polls, usability testing, and instrumentation of tools to help resolve the underlying research issues of "digital library vs. Web", "digital library vs. traditional library", "active learning vs. traditional". Transaction polls will be conducted at the end of episodes to determine how well users have achieved objectives with which they started an episode. Usability testing will be done to determine the effectiveness of the user interface by having users think aloud when performing a script of tasks, tasks they would normally encounter in the process of learning and/or preparing a course. Instrumentation of tools has the objective of gathering raw data as to the actual use of the environment, for example, describing what exists in the environment, what is used, how often it is used.

Below we outline the metrics we intend to use to understand the issues involved in the design of our digital library.

3.3.1 Impact on Learning

Preliminary validation of the hypothesis requires measurement of the impact of the digital library on the course structure and on the students' learning.

We shall describe a course by its objects such as syllabus, objectives, skills, software, notes, tests, and references; although it should be clear that the structure and number of these objects is itself an issue. Similarly, the actual learning experience of a student will be represented by objects such as homework, study material, term paper, exams, and portfolio. In the comparative study we need to measure the cost and benefit of using a digital library for : publication management (#metadata, %correct_metadata, time_to_create), resource location (precision, recall, response_time), archival value (#course_objects, #changes, #access, access_interval). For the active learning component of the evaluation study we need to measure the type and frequency of interactions and activities (#cbrowse, #NCSTRL+_access, #course_object_access) and the final outcome associated with these activities (exam_results, #skills_used_other_courses).

3.3.2 Impact on Delivery

Here we describe some basic metrics about the structure of objects, the overhead associated with their creation, and the validity of the resulting data to draw what conclusions. Specifically we intend to instrument NCSTRL+ to provide basic metrics for: bucket structure (#mime_elements/package/bucket, #elements/package/bucket, #course_objects_type, #learning_objects_type), learning activities (#search/cluster,#searches/keyword, #terms_condition_executed, #mime_elements_accessed), bucket creation (time_to_create/object). We will instrument the collaborative to provide the following basic metrics: #sessions, #participants/session, length_of_session, distance_participants/session.

D. References Cited

References

[1] ACM SIGIR on-line conference proceedings. http://turing.acm.org:8071.

[2] T. Bray, J. Paoli, and C. Sperberg-McQueen. Extensible markup language (XML) 1.0. Technical Report Technical Report REC-xml–19980210, W3C, Feb. 1998. http://www.w3.org/TR/REC-xml.

[3] D. Choy, C. Dwork, J. Lotspiech, L. Anderson, S. Boyer, R. Dievendorff, T. Griffin, B. Hoenig, M. Jackson, W. Kaka, J. McCrossin, A. Miller, R. Morris, and N. Pass. A digital library system for periodicals distribution. In *Proceedings of the Third Forum on Research and Technology Advances in Digital Libraries - ADL '96*, pages 95–103, 1996.

[4] R. Daniel and C. Lagoze. Distributed active relationships in the warwick framework. In *Proceedings of the 2nd IEEE Metadata Conference*, pages 16–17, Sept. 1997.

[5] J. R. Davis, D. B. Krafft, and C. Lagoze. *Dienst: Building a Production Technical Report Server*, pages 211–222. Springer-Verlag, 1995.

[6] J. R. Davis and C. Lagoze. A protocol and server for a distributed digital technical report library. Technical Report Technical Report, TR94-1418, Cornell University, 1994.

[7] J. R. Davis and C. Lagoze. The networked computer science technical report library. Technical Report Technical Report TR96-1595, Cornell University Computer Science, July 1996.

[8] T. Duffy and D. Cunningham. Contructivism: Implications for the design and delivery of instruction. In D. Jonassen, editor, *Handbook of Research on Educational Communication and Technology*. New York: Scholastic, 1995.

[9] D. Eckhardt, Jr., M. Jipping, C. Wild, S. Zeil, and C. Roberts. Open environments to support systems engineering tool integration: A study using the Portable Common Tool Environment. Technical Report Technical Memorandum 4489, NASA, Sept. 1993.

[10] EMSL Web-Tour. http://www.emsl.pnl.gov:2080/docs/collab/webarch.html.

[11] S. Esler and M. Nelson. Evolution of scientific and technical information dissemination. *Journal of the American Society of Information Science*, 49(1):82–91, 1998.

[12] E. Fox, J. Eaton, G. McMillan, N. Kipp, L. Wiess, E. Arce, and S. Guyer. National digital library of theses and dissertations: A scalable and sustainable approach to unlock university resources. *D-Lib Magazine*, Sept. 1996. http://www.dlib.org/dlib/september96/theses/09fox.html.

[13] E. Fox and L. Kieffer. Multimedia curricula, courses and knowledge modules. *ACM Computing Surveys*, 27(4):549–551, Dec. 1995.

[14] E. A. Fox and N. D. Barnette. Improving education through a computer science digital library with three types of WWW servers. In *Proceedings of the Second International WWW Conference*, pages 17–20, Oct. 1994.

[15] High/Scope. http://biz.map.com/~ecdgroup/hspc.html.

[16] IBM digital library version 2. http://www.software.ibm.com/is/dig-lib/v2factsheet/.

[17] R. Kahn and R. Wilensky. A framework for distributed digital object services. Technical Report cnri.dlib/tn95-01, CNRI, May 1995. http://www.cnri.reston.va.us/home/cstr/arch/k-w.html.

[18] C. Lagoze, C. A. Lynch, and R. D. Jr. The warwick framework: A container architecture for aggregating sets of metadata. Technical Report Technical Report TR96-1593, Cornell Computer Science, June 1996.

[19] O. Lassila and R. R. Swick. Resource description framework (RDF) model and syntax. Technical Report Technical Report WD-rdf-syntax-19980216, W3C, Feb. 1998. http://www.w3.org/TR/WD-rdf-syntax/.

[20] H. Lowe, W. Walker, S. Polonkey, F. Jiang, J. Vries, and A. McCray. The image engine HPCC project. a medical digital library system using agent-based technology to create an integrated view of the electronic medical record. In *Proceedings of the Third Forum on Research and Technology Advances in Digital Libraries - ADL '96*, pages 45–56, 1996.

[21] K. Maly, H. Abdel-Wahab, C. Overstreet, C. Wild, A. Gupta, A. Youssef, E. Stoica, and E. Al-Shaer. Interactive distance learning and training over intranets. *IEEE Internet Computing*, 1(1):60–71, Jan. 1997.

[22] M. J. Masullo. Putting digital libraries to work for education. In *OECD-KISDI Joint Conference on Information Infrastructure*, Apr. 1995.

[23] M. J. Masullo and K.-T. Huang. The EduPort demonstration project, a digital library system for K-12 education. In *Proceedings of the East-West Conference on Computer Technologies in Education*, Sept. 1994.

[24] S. McCanne and V. Jacobson. vic: A flexible framework framework for packet video. In *ACM Multimedia '95*, 1995.

[25] Microsoft NetMeeting. http://www.microsoft.com/netmeeting/.

[26] NCSTRL participating organizations. http://www.ncstrl.org/Dienst/UI/2.0/ListPublishers.

[27] M. L. Nelson and S. L. Esler. TRSkit: A simple digital library toolkit. *Journal of Internet Cataloging*, 1(2):41–55, 1997.

[28] M. L. Nelson, K. Maly, S. N. T. Shen, and M. Zubair. Buckets: Aggregative, intelligent agents for publishing. Technical Report CS TR-97-41, Old Dominion Univ., Dec. 1997.

[29] M. L. Nelson, K. Maly, S. N. T. Shen, and M. Zubair. NCSTRL+: Adding multi-discipline and multi-genre support to the dienst protocol using clusters and buckets. In *Proceedings of IEEE Advances in Digital Libraries 98*, pages 128–136, Apr. 1998.

[30] G. D. Ofiesh. Redefining the role of the teacher in the digital age. In *Teacher Education in the Use of Technology-Based Learning Systems Proceedings*, pages 48–53, 1995.

[31] PictureTalk. http://www.picturetalk.com/.

[32] PlaceWare. http://www.placeware.com/.

[33] M. Rees, R. Iannella, A. Lee, G. Smith, and T. Woo. Spinning a yarn: User interfaces for synchronous remote electronic meetings. In *Proceedings of OZCHI93*, pages 42–58, 1993.

[34] B. Schatz and H. Chen. Building large-scale digital libraries. *IEEE Computer*, 29(5):22–26, 1996.

[35] B. Schatz, W. H. Mischo, T. W. Cole, J. B. Hardin, A. P. Bishop, and H. Chen. Federating diverse collections of scientific literature. *IEEE Computer*, 29(5):28–36, 1996.

[36] TANGO. http://trurl.npac.syr.edu/tango/.

[37] The National Digital Library Program. http://lcweb2.loc.gov/ammem/ndlpedu/.

[38] The Virtual Hospital (VH). http://mulwala.looksmart.com:8080/?comefrom=ns1&divert.

[39] UVa SEAS Electronic Undergraduate Thesis Pilot. http://univac.cs.virginia.edu:3066/SEAS_ETD.html.

[40] S. Zeil and E. Epp. Interpretation in a tool-fragment environment. In *10th International Conference on Software Engineering*. IEEE CS, Apr. 1988.