Harvard University Library
Library Digital Initiative

Frances Loeb Library
Graduate School of Design

Web-based course material archiving project

Final Report
July 19, 2004
(Revised May 19, 2005)

Table of Contents

Introduction

This is the final project report for a Library Digital Initiative (LDI) grant received by the Loeb Design Library in 2001. After a year of inactivity, the Graduate School of Design (GSD) LDI Web-based course material archiving project officially restarted in November 2002 chaired by Kevin Lau. Some of the project timelines were adjusted to accommodate the new start date, but the original project goals, namely, to conduct a one year assessment of the needs and challenges of archiving web-based course materials with a primary focus on data from Fall 1999 to Spring 2003 held within a local learning management system (LMS) at the GSD, have not changed. As part of the study, the grant was to also address how identified problems and questions affected the University community as a whole.

The results after this one year preliminary study focused on identifying the key challenges of archiving web-based course materials, framing the right questions and issues to be addressed before the actual implementation phase, speculating on an array of possible solutions and finally sharing the findings with the Harvard community in the form of a University report and an open forum presentation. We are pleased to report that the goals as initially outlined have been met and indeed, surpassed.

Project Committees

To ensure a collaborative working environment and the inclusion of as much university expertise as possible, two project committees were formed; one to address policy matters and the second to work on technical issues. While the functional advisory committee addressed the administrative issues that affect the acquisition and preservation of web-based course materials, the technical advisory committee worked with the GSD's Head of Instructional Technology to come up with working models for preserving current-use and legacy course materials.

Both advisory committees came up with recommendations for establishing a University-wide model for collecting, preserving, and maintaining access to web-based course materials. They concluded that web-based course curricula represent a significant investment of University resources and play an integral part in the educational pedagogies of the University. Thus it is imperative for the University to address the long-term preservation of these materials. Furthermore, the committees concluded that such preservation should support the availability of these assets for both ongoing use in teaching and long-term research use.

In support of this goal, the committee members recommended the following:

- Examining the University's rights and interests in web-based course materials to preserve selected teaching materials, making these materials available for educational use, and retaining the non-exclusive right to copy or reformat the materials for the purposes of preservation and access.
- Establishing university policies to address the confidentiality of student information and students' intellectual property rights in course web sites.
- Developing appraisal guidelines for the selection of web-based course curricula.
- Developing archiving functions for each of the courseware platforms created at Harvard.

Finally, the committee members recommended the establishment of two new committees, one to address the legal and administrative challenges posed by the preservation of web-based course curricula, and the second to address the technical challenges in further depth. The membership of both committees, they decided, should represent University-wide constituencies and move beyond the goals of this study phase.

Project Participants

The members of the two advisory committees included specialists from the following schools:

Graduate School of Design, Frances Loeb Library
Kevin Lau, (Project manager)
Harvard Business School, Baker Library
Hal Bloom
Harvard College Library, Judaica Division
Violet Gilboa

Harvard Law School Library
David Warrington
Harvard Medical School, Countway Library of Medicine
Kathryn Baker
Harvard University Archives
David Best
Skip Kendall
Robin McEleheny
Harvard University Library, Office for Information Systems
Sue Kriegsman, (LDI MAP assistant)
Stephen Abrams
Robin Wendler
Harvard University Library, Weissman Preservation Center
Steve Chapman
Harvard University iCommons Group
Randy Myrick

Project Background

In 2001, the Francis Loeb Library at the Graduate School of Design received LDI funding for the study phase of the web-based course material archiving project. During this project the Loeb Library planned procedures and a technical infrastructure model for the documentation and archival of digital course materials in various formats. Two advisory committees surveyed file formats and software applications used to create course web sites and investigated copyright issues, documentation standards, and use and access models. The project also involved the design of an initial framework for addressing the implementation needs and method to support the capture and preservation of course web sites.

Two categories of course materials were studied in this project:

1. Legacy course web sites dating from Fall 1999 through Spring 2003 that were saved on storage systems.
2. Current-use course materials created in a variety of platforms.

Additionally, two audiences were identified for archival course materials:

1. Faculty and students seeking to continue to use a particular course curriculum in its original or updated form.
2. Researchers studying the development of curricula, teaching methods, and academic disciplines at Harvard.

The technical advisory committee worked with the GSD's Head of Instructional Technology to formulate working models for preserving the legacy and current-use course materials, while the functional advisory committee addressed the administrative and policy issues that affect the acquisition and preservation of web-based course materials.

As the result of this project, the advisory committees have made a number of recommendations in support of a University-wide model for collecting, preserving, and maintaining access to web-based course materials.

The Importance of Course Materials

As its primary purpose, an academic archive documents and preserves the corporate memory of the institution, including the activities and decisions it undertakes to carry out its mission: the education of students, the creation of new knowledge through research, and the sharing of that knowledge for the benefit of the institutional community and society at large.

In an academic organization, corporate memory serves a number of purposes. A clear view of the past not only provides a foundation and context for future decisions, but also chronicles the accomplishments of the institution and its constituents, maintains the identity of the institution, documents the role of the institution in the development of higher education and the progress of the professions, and supports the long-term value of graduates' credentials.

Academic endeavor – teaching and research by faculty and students – distinguishes universities from other types of institutions and distinguishes one university from its peers. Thus the core of any university archives must be the record of the courses taught in the university's degree programs.

Researchers working in such fields as intellectual history, the history of higher education, and the rise of the professions have long appreciated the importance of syllabi, lecture notes, students' class notes, and other artifacts of teaching as a means of providing historical "snapshots" of the development of a discipline. The survival of this type of material, however, has often been haphazard. Fortunate, indeed, are those educational institutions and professional associations that have systematically collected the evidence of what was being taught to its students and neophyte practitioners and the methods and successes of this pedagogy.

The advent of web-based courses has created new audiences for curriculum materials – faculty and students who want to use or refer to the materials over the long term with the convenience of round-the-clock remote access.

Course Materials at Harvard

Up to the present, Harvard has documented teaching and curricula in a variety of ways to serve a variety of purposes, with the scope of course documentation depending on the academic culture and administrative resources at each school. Some of the resulting documentation makes its way into the University's archives, but much of it remains in active department files or is discarded once the immediate need for it has been met.

Contributing to the heterogeneous nature of the documentation of course materials is the fact that archivists at Harvard have not defined teaching materials as University records. Thus these materials have not always been a focus of archival collecting efforts.

Harvard's schools now emphasize and encourage the use of information technology to expand the classroom experience and facilitate communication among faculty and students. Course catalogs, for instance, are available on the web and are fully searchable. Course web sites are increasingly common as well, and several schools have further established instructional technology programs to support the development and use of electronic teaching resources.

At the Graduate School of Design, for example, faculty members tend to use web-based curriculum materials developed for a particular course as building blocks for new materials. The school's portal for teaching resources has also encouraged the sharing, re-use, and further development of course materials.

Network-based teaching materials now represent a significant investment of the University's resources, from network upgrades and hardware purchases to labor on the part of IT specialists, teaching fellows, and faculty. Early experience shows, however, that web-based materials are some of the University's most evanescent assets, having been created in a wide variety of formats and with storage and maintenance handled locally on a case-by-case basis. The establishment of the Harvard Academic Computing Committee and the subsequent emergence of iCommons offer new opportunities to capture and preserve the academic experience in a systematic way. It would be ideal for the University to take advantage of the opportunity to proactively roll in collecting, preserving, and maintaining access to these materials in an even more systematic and comprehensive manner.

Challenges

Three types of challenges dominate any attempt to preserve web-based teaching materials: technical, administrative, and legal. The University must address these challenges and develop long-term solutions before it can embark on a program to collect and preserve these materials.

Technical challenges

These involve the capture, storage, and preservation of course web sites, as well as the development of procedures for making the materials accessible and functional over the long term. Technical issues will vary, depending on whether archivists and administrators are dealing with legacy teaching materials or current-use materials.

With the coordinated development and expanding implementation of the iCommons courseware platform at the University, the establishment of systematic methods for long-term preservation and access is more likely to be achieved. It should be noted, however, that some formats will fall outside the scope of established models and will require assessment on a case-by-case basis.

Administrative challenges

These involve decisions regarding the assignment of responsibility for web-based teaching materials both in current-use and in long-term research environments. Some course web sites may be best preserved and distributed centrally, in one of the University's archives. Other web-based curriculum materials may be better preserved and made accessible to researchers from one of the archives while leaving the relevant department to maintain the same materials locally for ongoing teaching purposes.

Administrative challenges also include the appraisal of web-based course materials in the context of short- and long-term University needs, as well as the allocation of resources to implement the appropriate preservation and access methodologies. In addition, administrative policies need to be established at the local and University levels to serve as guidelines before any implementation process can occur.

Legal challenges

These concern intellectual property rights to web-based teaching materials, and privacy rights governing student information that may appear in course web sites. While archivists, technical specialists, and administrators can undoubtedly resolve the technical and logistical issues, policies governing the legal issues must be established before the curriculum materials can be made available for use by the University community and outside researchers.

According to current University policy, members of the faculty retain the rights to their intellectual output. The iCommons courseware platform has been developed with the understanding that the University owns the courseware itself, while the faculty who use it own the content that they have created. Before the University's archives can devote any resources to acquiring web-based course materials, the right to preserve and maintain these materials must be clarified.

The committee members recommended the establishment of a University policy that will give the University the non-exclusive right to collect and preserve those teaching materials created at the University and to make them available for educational use, including the non-exclusive right to copy or reformat the materials for the purposes of preservation and access.

As an alternative but less-desirable option, committee members also recommended the creation of a standard agreement, covering the same rights described above, that each school can distribute to those members of its faculty and other parties who create online course materials and own the intellectual property rights.

Even if copyright is not in question, however, some faculty members may not feel comfortable with sharing their work with peers or students from other institutions. Therefore, in addition to addressing questions of copyright, it is possible that the option of restricted access to course materials will actually encourage some potential faculty and other contributors to make donations to the archive. The University PIN

authentication system could potentially be utilized to help restrict access to experimental or personal faculty research materials to a designated Harvard community.

Web-based course materials may not be limited to the work of faculty, however. These materials may also contain bulletin boards, threaded discussions, course email distributions, and project materials created during the duration of a course by students and guest lecturers – intellectual product ownership of which is generally considered to belong to the participants. A definitive University policy will be needed to address these issues.

Additional considerations (FERPA, appraisal)

FERPA, the Family Educational Rights and Privacy Act (20 U.S.C. § 1232g; 34 CFR Part 99), is a federally mandated law that protects the privacy of student education records. This Federal law applies to all schools that receive funds under an applicable program of the U.S. Department of Education. A set of guidelines on student privacy levels has been provided by the Office of the Provost to each school's registrar's office.[1] For instance, the Graduate School of Design's policy requires all students to complete the Directory Information section of the GSD Student Data form during both the fall and spring registration process. Students are given the opportunity to select their preferred privacy level concerning the use and release of their own data. If a student does not select one of the options listed on the form, the school's selected privacy defaults, based on the University recommendations, are applied. All data, including changes, are fed to the University LDAP system and also retained locally within the school's administrative database, which serves as the primary data source for authentication, access, and all administrative back-end process for the GSD's Courseware learning management system. These FERPA flags, currently in effect, within the administrative database allow the school to use a variety of XSLT filters to exclude or specifically suppress any student data so tagged from potentially being archived in future projects.

> Example from the GSD Courseware archival project of what a tei:name tag for a student who has chosen not to have any information available would look like:
> <tei:name type="person" key="false" gsd_cw:huid="false">false</tei:name>

Most, if not all, of the learning management system platforms in use at Harvard are connected with administrative and registrarial information systems at various other schools. Much of the information provided by these systems is subject to the FERPA and therefore consideration must be given to the confidentiality of student information that may be included in course web sites and guidelines needs to be in place to help archivist remove student information protected by FERPA from web sites as part of the process of preparing the course materials for archival acquisition and processing.

---

[1] The Office of the Provost's privacy level recommendations:

Privacy Levels

1) (Restricted) – Most restrictive; share only as mandated by law or faculty policy.
2) (Department Level Only) – Publish only within the School in online or printed format. Individual departments must not publish in online locations that can be accessed by people outside the School.
3) (Internal Harvard Printed Documents) – Publish only in internal Harvard printed documents, and provide to the Harvard Telephone Operators, Internal printed documents include the Student Telephone Directory. Note that these documents may be available to the public in certain University locations. Telephone operators may provide email address or campus mail center information to callers.
4) (Inside Harvard) – In addition to Level 3 publication, accessible electronically to users of computers on University data networks; or to users of computers outside of Harvard if they identify themselves by using the University PIN system.
5) (Public) – Accessible electronically to anyone on the Internet.

Project Web Site and Open Forum Presentation

As part of the process to help share the findings of this project between the Harvard community and the committee members, a web site was created to facilitate communicate and provide a central portal to distribute information.
The Uniform Resource Locator (URL) to the web site is: http://hul.harvard.edu/ois/projects/gsdarchiving/

Within this site, all the meeting minutes for both committee groups have been summarized to provide an insight into the development process throughout the project's life-cycle. In addition, the three presentations used in the well attended open forum presentation on March 19, 2005 at the Lamont Library Forum room can also be found here.
The open forum presentation topics include:

- LDI project summary and the GSD case study (Kevin Lau)
  http://hul.harvard.edu/ois/projects/gsdarchiving/LDI_brownbag_GSD_case_study.pdf
- The Report to the University & Current policy for the archives and records management in relation to the report (Robin McEleheny & David Best)
  http://hul.harvard.edu/ois/projects/gsdarchiving/LDI_brownbag_University_Report.pdf
- The technology implications and concerns raised by the Report to the University (Stephen Abrams)
  http://hul.harvard.edu/ois/projects/gsdarchiving/LDI_brownbag_Technology_Implications.pdf

Project Staff

Kevin Lau, Project manager	Graduate School of Design
Marc Moskowitz, Project programmer	Graduate School of Design
Sue Kriegsman, LDI MAP assistant	Harvard University Library, Office for Information Systems

Budget

The total awarded funding cap for this LDI project is $64,000. The final grand total expenses (through June 30, 2004) are $50,129.95.

Funds have been used exclusively for a temporary .5 FTE programmer for 8 months and a .2 FTE LDI MAP assistant for 20 months. The project programmer focused on the inventory of the course materials/data, investigating archival methodologies, and developing/reviewing XML metadata schemas. The LDI MAP assistant contributed many of the administrative duties, project documentation, coordinating project meetings, and providing varies systems/project support.

Cost of Web-based Course Material Archiving Project (Study Phase)

Year one	LDI Funding	Cost Share
Salaries (incl. Benefits)		
Project manager @ .25 FTE	$0.00	Cost share
Project programmer @ .5 FTE	LDI funds	$0.00
LDI MAP assistant @ .2 FTE	LDI funds	$0.00
Equipment (incl. hardware/software)		
Workstation for project programmer	$0.00	$2,250.00
Data storage/temporary local archival server space	$0.00	$5,000.00
Software and supplies	$0.00	$2,000.00
Total Funds	$50,000	$31,000

Graduate School of Design Case Study

Acronyms

CAD	Computer-aided design
DRS	Digital Repository Service
DTD	Document Type Definition
GIF	Graphics Interchange Format
GIS	Geographic Information System
GSD	Graduate School of Design
HTML	Hypertext Markup Language
IT	Information Technology
JPEG	Joint Photographic Experts Group
LMS	Learning Management System
METS	Metadata Encoding & Transmission Standard
Perl	Practical Extraction and Report Language
SCORM	Sharable Content Object Reference Model
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TEI	Text Encoding Initiative
TIFF	Tag Image File Format
XML	Extensible Markup Language

GSD Policy Decisions

Included here are some key points identified during the GSD case study that will be shared locally and could be applied to other digital materials as well. Addressing and defining some of these points prior to any implementation process will help the GSD fulfill its archival goals and move toward a proactive approach to preserving its history of web-based course materials rather than reactively dealing with specific issues as they present themselves.

Important factors for the GSD to consider:

1) Define the school's archival mission and goals.
2) Identify key players, stakeholders, and local resource providers. Know what service they can and cannot provide.
3) Compile an inventory of information to preserve.
    a. Types of information:
        i. Faculty-created material
        ii. Student-created material
        iii. Co-mingled student and faculty material
        iv. Externally owned material (copyrighted material like journal articles or excerpts from books that may be part of class reserve readings)
        v. Administrative and management data (enrollment information, interactive quizzes, access/usage logs, etc.)
        vi. Other miscellaneous institutionally generated materials (librarian-created guides or documentation on computing facilities prepared by local IT group)
    b. File format types (example from the Design School):
        i. text file types, including HTML
        ii. still image file types
        iii. CAD (computer-aided design) file types

iv. GIS (Geographic Information Systems) file types
v. multimedia file types

c. Administrative metadata:
i. Content ownership
ii. Author(s)
iii. Creation dates
iv. Content descriptions

d. Technical metadata
i. File format types
ii. Applications used
iii. Version of the applications
iv. Size, orientations, and resolutions
v. Interrelationships

e. Use and appraisal metadata:
i. Importance of content
ii. Retention periods
iii. Access permissions/audience

f. Storage and location of metadata
(image of CPU, backup tapes, floppies, hard drive)

4) Develop appropriate high-level organizational policies and objectives, ensuring that they are clear and well communicated.
   a. Because there are many types of information and formats, established policies can serve as a preservation framework for decision-making choices over time, particularly for questions pertaining to ownership and use of student/faculty work.
   b. "The primary goals of policy making would be to protect and empower the faculty and the institution by making sure that questions involving privacy and rights are managed appropriately." - Clifford Lynch, Fall 2002 CNI Project Briefing
5) Address intellectual rights/copyright issues.
   a. Student- and faculty-created material
   b. Externally owned material
6) Develop appraisal and retention protocols, and set a time frame for determining when information becomes frozen as a record vs. remaining a living document.
   a. Systems to house information
   b. Systems to permit access during period as living document
   c. Recovery and backup strategies
7) Address access rights to information and under what circumstances access would be granted.
   a. Post-course faculty access
   b. Post-course student access
   c. Re-use of courses
   d. Post-course independent access
8) Identify and categorize information for long-term preservation and information in active record-keeping systems that might only need to be kept for a relatively short time-horizon along with information that has no long-term historical value for the organization.
9) Develop a process to catalog the information thoroughly, giving particular importance to its provenance and transformations of the material.
10) For a given digital resource, address whether its original presentation form (the literal content) or only its intellectual content must be preserved.
   a. For example, it is common at the GSD for information related to a single project to exist partly in traditional paper format, partly in an email box, word-processing file, database tables, and multiple image files generated in a variety of formats, platforms, and programs.
   b. Because of the hybrid nature of information produced, one challenge is to design a suitable solution for the school.
      For example, Geographic Information System (GIS) produced maps, CAD and 3-D images converted and flatted into a two-dimensional image may not be suitable

substitutes for the originals. Converted files may lack elements or other critical information vital to the original form.

11) Address organizational responsibilities, ongoing maintenance, and resources.
Because maintaining archived materials over time will require ongoing investment and attention, addressing economics could very likely further narrow what resources would be deemed as valuable and identified for long-term preservation.

Scope of Courseware System and Estimated Resources Required

- Courses examined (Fall 1999 through Spring 2003): 1443
- Size of generated course data: 8298510 bytes (7M 936K 14B)
- Uploaded files: 4836
- Edited application files: 48
- Courseware modules/tools in original application: 11
- Time commitments (estimated):
  - Learning existing system infrastructure: 4 weeks
  - Learning new technologies: 2 weeks
  - Learning about possible solutions for data storage: 4 weeks
  - Editing the front end of existing application: 15 weeks
  - Writing code to access and manipulate generated data: 9 weeks
  - Documentation: 4 weeks
- Skill set:
  - Significant knowledge of (or willingness to learn) language in which the application is written. In the case at the GSD, this is Cold Fusion.
  - SQL, as part of work with the existing application. In theory, this would not be necessary to alter an application that made no use of SQL.
  - XML and HTML, to convert an existing HTML front-end into XML.
  - PERL, for use in accessing and manipulating data generated by the web site. Similar work could have been done in another general-purpose scripting language, such as Python.
  - XSLT, used to transform XML data.

Case Study Introduction

One goal of the Graduate School of Design is to document its own educational process. With an increase in electronic teaching tools, it became apparent that a plan to capture and archive web-based course materials would need to be established in order to collect current materials. The purpose of the Library Digital Initiative (LDI) grant was to study the scope of the course materials and the options for capturing and archiving the online web-based content.

In 1999 the Graduate School of Design developed Courseware, an in-house Learning Management System that served as a single entry portal to various pedagogical tools that provided GSD faculty with new ways to teach and deliver course materials. The Courseware architecture was designed with a back-end connection heavily dependent on the school's administrative Oracle database to gather student, course, and other administrative information to determine authentication, set levels of access, and integrate administrative data within course web sites. Essentially, the Courseware system worked as a Web-based learning system, delivering content unidirectionally from the server to the user via the Internet, with data dynamically generated and access determined by the GSD Registrar's database. Occasionally, files would be uploaded by users or information would be posted back to the server through specific tools such as discussion threads that further supplemented the content of the course. Because of the close interaction of the Courseware LMS with the administrative database, a reference trail was logged for all files uploaded to the associated course. Specific course content would occasionally be used as a building block for the course offered in a different term. But by fall 2002, due to changes in the school's direction, development of the

Courseware system was put on permanent hold as the school contemplated switching to the University's iCommons LMS, however, the GSD continued to use Courseware through the spring 2003 term.

In the LDI project, technical specialists at the GSD surveyed materials in the Courseware system dating from fall 1999 through spring 2003 in order to assess the range of formats included and the possibilities for their long-term preservation. Given the large variety of formats used in many GSD courses and the school's objective of long-term preservation, many difficult decisions had to be made. With limited resources available in the GSD, the project relied on the services and expertise currently available to the Harvard community through the Library's Office of Information Systems, University Archives, and on similar work done outside the University rather then trying to develop new resources or methodologies. Thus the proposed preservation model involves use of the LDI Digital Repository Service (DRS) for long-term storage of digital files and a form of Extensible Markup Language (XML) encoding schema for metadata that would be readily usable, retain all course information, and be customizable in order to include reference tags for associated files not housed in the database was determined to be the most logical format for storing the dynamically generated data for each class. The TEI-lite Document Type Definition (DTD) was chosen as a metadata schema to identify and describe the digital objects in the system and their interrelationships.

In the course of this project, technical specialists were able to create a model that could allow for the long-term preservation of the materials in the Courseware LMS. Full implementation of this model, however, will require a number of policy and case-specific decisions. Project participants drew the following conclusions from the GSD case study:

- In order for the archival of LMS materials to be scalable and fully successful, administrative policies and technical specifications designed to support preservation and long-term access must be established at the outset of system design.
- Maintaining appropriate metadata is vital for ensuring the durability of valuable digital resources, regardless of the DTD standard or model used.
- It will be necessary to involve trained archivists, metadata analysts, and technical specialists in the early architectural design stages of system/application development.

GSD Technical Challenges and Solutions

Because the Courseware system is so intimately interwoven with the school's administrative database, and bearing in mind the issues involved with intellectual property rights, the technical assessment of the Courseware system began from the perspective of how to best utilize the existing references provided by the administrative database, as opposed to examining all of the individual files uploaded to the system and then determining those files' respective course associations. This process enabled materials not related to a given course to be sorted out.

A form of XML that would be readily usable, retain all course information, and be customizable in order to include tags specifically referencing associated files not housed in the database was determined to be the most logical format for storing the dynamically generated data for each class. One option would have been to create an in-house DTD (document type definition) specific to the project that mimics the structure of the Courseware records for a class. Since this approach requires a complex structure to be maintained indefinitely, the decision was made to choose instead an open-source structure as a base that can be elaborated while still allowing liberal use of type and identification attributes to distinguish various Courseware modules and their sub-elements, including a namespace to refer to attached files as well as an

attribute to refer to users by their Harvard University ID numbers.[2] The low-complexity DTD, TEI-Lite, from the Text Encoding Initiative Consortium (an independent organization supporting "an international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent") was selected as the foundational building block for this initial assessment.

During the final research/revision phase of the technical assessment process, several other standards were brought to our attention, most notably METS (Metadata Encoding & Transmission Standard) and SCORM (Sharable Content Object Reference Model), which may well provide a better fit for these data. In a nutshell, METS is "a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library." It is based on the structure of a digital work and provides a useful way to record information about the content of the work, especially about uploaded files. In contrast, "at its simplest," SCORM is "a model that references a set of interrelated technical specifications and guidelines designed to meet [the Department of Defense]'s high-level requirements for Web-based learning content." It is unclear based on our current assessment whether the content and metadata available in the Courseware system are compatible with the goals and requirements of SCORM. A significant goal of the Advanced Distributed Learning initiative, which produced the SCORM, is to enable ubiquitous availability of individualized instruction. A more in-depth inventory of Courseware's content, with a specific focus on sharable educational content, is probably necessary for true conformance with SCORM. Creating an XSLT document that, given our existing TEI-based XML documents, can render new XML files which conform as much as possible to one or both of these standards, would be feasible given enough time and resource.

However, this brings up an important point. The Courseware system was designed for the convenience of professors and students in a face-to-face academic environment and was not optimized for long-term storage. Most notably, there is significant metadata absent from the Courseware content because there was no developed infrastructure to collect various metadata when the content was created. The cost of leaving out this information is significant. In order to store every file housed in the Courseware system with the name and version of the application used to create it, it would be necessary for someone to manually go through all 2,774 files in the system, track down the creator of each file, and find out which version of the application that person was using at the time that that file was created. In many cases, this information would probably be lost permanently and so important metadata fields would be left vacant. Conversely, if metadata considerations had been taken into account when the system was initially designed, the requested

---

[2] In order to make this back-and-forth conversion possible, we have taken the following steps have been taken:

1) Made liberal use of type and id attributes within TEI-Lite to distinguish various CourseWare modules and their sub-elements.
2) Created a separate namespace "gsd_cw" in our generated documents, that tags information specific to this project. This namespace includes a tag to refer to attached files, as well as an attribute to refer to a tei:person by Harvard University ID number.

Examples:
Complex structure:
<gsd_cw:attach ref='/2003/spring/1212/cw/cw_mail/BDEEC839-2EB4-11D7-976A00D0B7741CA9/group_mappings_pinup_2=7=03.doc?823571964323393848'>group_mappings_pinup_2=7=03.doc <tei:date value="2003-02-05 8:59 PM">2003-02-05 08:59 PM</tei:date></gsd_cw:attach>

Simple structure:
<gsd_cw:attach ref="2003/spring/1404/cw/012903_lecture.pdf" />

(See appendix for complete sample xml and xsl stylesheet for a test course)

metadata would now be available, but possibly at the cost of reduced participation due to the user inconvenience involved in needing to provide this sort of information. The decision here involves the familiar tradeoff between extra effort early in a process and increased cost later. In this case, the cost could well be a sacrifice in the long-term availability of the data.

It is our understanding, which may be incomplete, that METS is focused on the general needs of a library managing collections of digital work. SCORM, however, has the very specific goal of sharing and aggregating content for possible use in a larger learning application.

Courseware is not the sort of application for which SCORM was designed. It is entirely reasonable that a future project may wish to make this sort of data available for aggregation, but the current project, based as it is on a curricular website that is supplemental to a classroom-based educational environment, is not particularly amenable to aggregation. While there are digital objects in the Courseware data and some of the data might be usable for distance learning, the data cannot stand alone outside the course since it only provides meaningful value when evaluated as part of the whole course curriculum.

METS may be a good choice to help further organize and supplement the limited metadata we have gathered, regardless of how many original element additions are included in the base DTD of TEI-Lite, as the point is well taken that more metadata for these objects is needed. However, much of the requested metadata is not part of the stored data in the system and may need to be inferred, especially since the development of the Courseware system is on permanent hold. We are aware that TEI is not an ideal solution and that a better one may exist for storage of the curricular website information, but in our limited time we did not find it. TEI is a free-form standard that allows us to replicate the various free-form structures of the Courseware data in as close a form as possible to their inherent structure. This specific solution is applicable only for the GSD Courseware data, as the system was designed for the convenience of professors and students in a face-to-face academic environment. It is therefore optimized neither for long-term storage nor for autonomous distance learning. The decision may be made in the future to require more detailed information from participants in similar systems, but that information will have a convenience cost that will either reduce participation or require tools to be built that will mitigate such a cost by providing the information in an automated way to assist creation.

Besides the various text file formats (such as .txt, .pdf, .doc), the school also has several types of image file formats to examine as well. [3] Images currently in acceptable DRS formats (such as TIFF, JPEG, and GIF) would be left as they are. During the inventory, a significant variety of file formats used for CAD (Computer-aided design) data were discovered. The various programs that read and write these formats are unlikely to all be available in the indefinite future. This led us to search for an independently-supported standard format in which to store this data. We believe that the SVG (Scalable Vector Graphics) format, an

[3] 11 different file media types encompassing 51 specific file formats were identified on the active web directories at the Graduate School of Design. The inventory took approximately 80 hours to complete.

Media type	*Format (by file extension)*
CAD	dcd, dwg, dxf, efc, max, mcd
3D	3ds, fmz, wrl
GIS	apr, aux, db, dbf, dir, met, mperr, mxd, shp, shx
Images	ai, bmp, eps, gif, jpg, pct, plt, png, psd, tif
Text	pdf, ppt, ps, qxd, rtf, txt, doc
Mark-up	css, htm, js, xml
Animation	fla, swf
Audio	au, mid, mp3, ram, wav
Video	avi, mov, mpg
Spreadsheet	xls
Compressed	zip

XML-based solution for storing visual data in a resolution-independent manner, is the best standard currently available for CAD data. For this process we chose an off-the-shelf application that converts and keeps some, though not all, of the associated metadata using XML, XSLT, and SVG with built-in layer controls and navigation tools, plus markup/redlining tools. As no conversion process currently available mimics every aspect of a complex file's original unique attributes, however, the storage of the original files, is still highly recommended so that a better solution can be adopted if one becomes available while the original files are still readable. However, for GIS (Geographic Information System) files, our initial test indicates that vital interactive layers were either rendered useless or lost in the conversion process to SVG, thus this might not be an ideal solution. Further investigation is required, and a separate assessment project to focus the proper time, resources, and research is recommended.

Lessons Learned

It is clear that maintaining appropriate metadata is vital in ensuring the durability of valuable digital resources, regardless of the DTD standard or model used. Detailed metadata would serve as the connective link and retrieval and management vehicle for all the components that make up a single course. In the case of the GSD, it was logical to take advantage of all existing administrative information and descriptive references that the database held in order to supplement core course information. For by gathering this comprehensive information now, we can always create filters later to selectively extract desired information for any future archival projects.

Keeping in mind the demands of metadata for archival purposes is critical during the design phase of any LMS application. Perhaps the most important lesson learned during this study phase is the need for established policies, whether they are University-wide and general or more local and specific, *before* any implementation begins. Established policies can serve as a preservation framework for decision-making choices over time, especially in addressing questions pertaining to intellectual property and for communicating archival expectations up front.

References

1. A joint digital repository project for University and MIT. *Cambridge University*. 2003 February 10. http://www.admin.cam.ac.uk/news/press/dpp/2003021001

2. Clifton, G., Day, M. What's new in digital preservation. *DPC/PADI (Digital Preservation Coalition and Preserving Access to Digital Information)*. 2003 October 27; June – September(6). http://www.dpconline.org/docs/whatsnew/issue6.html

3. *DSpace*. MIT Libraries & Hewlett-Packard Company. http://dspace.org/

4. *DSpace@Cambridge*. Cambridge University Library and the MIT Libraries. http://www.lib.cam.ac.uk/dspace/

5. *Harvard University General Records Schedule*. The President and Fellows of Harvard College. 2003 December 17. http://grs.harvard.edu/

6. *JHOVE (JSTOR/Harvard Object Validation Environment: a project to develop an extensible framework for format validation)*. http://hul.harvard.edu/jhove/

7. *JISC (The Joint Information Systems Committee Digital Preservation site).* http://www.jisc.ac.uk/index.cfm?name=pres_home

8. Lynch, C. The Afterlives of Courses on the Network: Information Management Issues for Learning Management Systems. *CNI (Coalition for Networked Information)*. 2003 February 11. http://www.cni.org/tfms/2002b.fall/abstracts/PB-Afterlives-Lynch.html

9. *METS (Metadata Encoding & Transmission Standard)*. http://www.loc.gov/standards/mets/

10. *NDIIPP (National Digital Information Infrastructure Preservation Program).* http://www.digitalpreservation.gov/

11. *NIST (National Institute of Standards and Technology) Data Preservation Program.* http://www.itl.nist.gov/div895/isis/datastorage.html

12. *Penn State-Internet2 to develop authenticated P2P software called LionShare*. Penn State University. 2003 October 12. http://its.psu.edu/news/LionShare.html

13. *SCORM (The Sharable Content Object Reference Model).* http://www.adlnet.org/scorm/index.cfm?CFID=73332&CFTOKEN=131c77d2ab7339af-090A01F5-ECB2-2D53-0275BC642D556929&jsessionid=54302eca9e97$07$B8$C

14. *TEI (Text Encoding Initiative)*. http://www.tei-c.org/

15. *The Technology, Education and Copyright Harmonization (TEACH) Act. ALA.* http://www.ala.org/Template.cfm?Section=Distance_Education_and_the_TEACH_Act&Template=/ContentManagement/ContentDisplay.cfm&ContentID=25939

Appendix

Example: TEI XML for GSD Test Course

```
<?xml version="1.0" encoding="UTF-8"?>
<tei:TEI.2 xmlns:tei="http://tei-c.org.uk/Lite/"
xmlns:gsd_cw="http://www.gsd.harvard.edu/library/information_systems/courseware_archive.html">
        <tei:teiHeader>
                <tei:fileDesc>
                        <tei:titleStmt>
                                <tei:title>101:Test Course: Independent Study Fall 2002:machine-readable
form of instructional website</tei:title>
                                <tei:author>
                                        <tei:name type="person" key="cwadmin@gsd.harvard.edu"
gsd_cw:huid="">instructor itg</tei:name>
                                </tei:author>
                                <tei:sponsor>Harvard Graduate School of Design</tei:sponsor>
                        </tei:titleStmt>
                        <tei:publicationStmt>
    Originally published on a web site of the Harvard Graduate School of Design.
    </tei:publicationStmt>
                        <tei:sourceDesc>
    </tei:sourceDesc>
                </tei:fileDesc>
        </tei:teiHeader>
        <tei:text>
                <tei:front><![CDATA[
]]></tei:front>
                <tei:body>
                        <!-- TOC items -->
                        <tei:div type="file">
                                <tei:head>Syllabus</tei:head>
                                <gsd_cw:attach ref="2002/fall/101/cw/syllabus.c"/>
                        </tei:div>
                        <tei:div id="announcements">
                                <tei:div type="announcement">
                                        <tei:head>testing jan 31 (<tei:date value="2003-01-10 4:00
PM">2003-01-10 04:00 PM</tei:date>-<tei:date value="2003-01-31 4:00 PM">2003-01-31 04:00 PM</tei:date>
                                                <tei:name type="person" key="
Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Student_Name</tei:name>)</tei:head><![CDATA[
<TT>testing</TT>
]]><tei:note type="priority">4</tei:note>
                                </tei:div>
                                <tei:div type="announcement">
                                        <tei:head>are annc&amp;apos;s fixed? (<tei:date value="2002-11-01 4:00
PM">2002-11-01 04:00 PM</tei:date>-<tei:date value="2003-11-01 4:00 PM">2003-11-01 04:00 PM</tei:date>
                                                <tei:name type="person" key="
Student_Name@gsd.harvard.edu" gsd_cw:huid="
HUID_Number">Student_Name</tei:name>)</tei:head><![CDATA[
<TT>this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's
fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are
annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? <BR><BR>this is a test; are annc's fixed?
this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's
fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? v</TT>
]]><tei:note type="priority">4</tei:note>
                                </tei:div>
                                <tei:div type="announcement">
                                        <tei:head>are annc&amp;apos;s with attch&amp;apos;s fixed? (<tei:date
value="2002-11-01 4:00 PM">2002-11-01 04:00 PM</tei:date>-<tei:date value="2003-11-01 4:00 PM">2003-11-01
04:00 PM</tei:date>
```

```
                                                <tei:name type="person" key="
Student_Name@gsd.harvard.edu" gsd_cw:huid="
HUID_Number">Student_Name</tei:name>)</tei:head><![CDATA[
<TT>this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? <BR><BR>this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? v</TT>
]]><tei:div type="attachments">
                                                <tei:head>Attachments</tei:head>
                                                <gsd_cw:attach
ref="/2002/fall/101/cw/cw_announcements/23F044DC-EC22-11D6-
976A00D0B7741CA9/internet_map.gif?033701666748065041">internet_map.gif</gsd_cw:attach>
                                        </tei:div>
                                        <tei:note type="priority">4</tei:note>
                                </tei:div>
                                <tei:div type="announcement">
                                        <tei:head>Test Announcement - no class today, 9/20th (<tei:date
value="2002-09-20 4:00 PM">2002-09-20 04:00 PM</tei:date>-<tei:date value="2002-09-21 4:00 PM">2002-09-21
04:00 PM</tei:date>
                                                <tei:name type="person"
key="cwadmin@gsd.harvard.edu" gsd_cw:huid="">instructor itg</tei:name>)</tei:head><![CDATA[
<TT></TT>
]]><tei:note type="priority">4</tei:note>
                                </tei:div>
                        </tei:div>
                        <tei:div id="assignments">
                                <tei:table rows="5" columns="5">
                                        <tei:row role="label">
                                                <tei:cell>Students</tei:cell>
                                                <tei:cell>
                                                        <tei:head>pop quiz</tei:head>
                                                        <tei:div type="description"><![CDATA[
<TT>here are the results of the pop quiz from September 15. </TT>
]]></tei:div>
                                                        <!-- projects & grades -->
                                                        <tei:note type="assigned">
                                                                <tei:date value="2002-09-13 12:00
AM">2002-09-13 12:00 AM</tei:date>
                                                        </tei:note>
                                                        <tei:note type="due">
                                                                <tei:date value="2002-09-15 10:00
AM">2002-09-15 10:00 AM</tei:date>
                                                        </tei:note>
                                                </tei:cell>
                                                <tei:cell>
                                                        <tei:head>Graph Searching</tei:head>
                                                        <tei:div type="description"><![CDATA[
<TT>You must write a total of four separate search programs: BREADTH-FIRST, DEPTH-FIRST, LEAST-COST, and ASTAR. However, if you plan your code well, you may find that the four are nearly identical in most respects, and in fact you may be able to write a single basic search function with four top-level function calls to it.<BR><BR>For your A* algorithm, use the air distances between two cities as the heuristic estimate for the remaining cost.</TT>
]]></tei:div>
                                                        <tei:div type="details">
                                                                <gsd_cw:attach
ref="http://www.people.fas.harvard.edu/~albert/cscie220/asst2.html"/>
                                                        </tei:div>
                                                        <!-- projects & grades -->
                                                        <tei:note type="assigned">
                                                                <tei:date value="2002-10-22 12:00
AM">2002-10-22 12:00 AM</tei:date>
                                                        </tei:note>
```

```
                                                <tei:note type="due">
                                                        <tei:date value="2002-10-22 8:00
AM">2002-10-22 08:00 AM</tei:date>
                                                </tei:note>
                                        </tei:cell>
                                        <tei:cell>
                                                <tei:head>Mimimax Searching</tei:head>
                                                <tei:div type="description"><![CDATA[
<TT>GET-MOVE and the associated functions you write must be an implementation of the minimax routine (alpha-
beta pruning is an optional extension). You will need to write a set of functions to properly implement this routine, but
GET-MOVE must be the top-level function and must work as stated above. If you take advantage of LISP's recursive
nature you will find that this problem is very simple to program. A pair of mutually recursive functions can perform
almost all of the work of maintaining the search tree for you, with no actual tree data structure
required.<BR><BR>You will need to write a function which "expands" a node, but notice that this task is fairly simple
since all you need to do is check which moves are legal and then run %DO-MOVE on each of them, giving it the
appropriate player (which will be different at each level of the tree), the hole number of the legal move, and the board
position of the node you are expanding. (Don't send %DO-MOVE an illegal move! It won't check, and anything might
happen!)<BR><BR>Since minimax relies on a heuristic evaluator to avoid searching all the way to the end of the
game, you will need to write such an evaluator. Unless you find yourself enchanted by the prospect, you needn't spend
too much time worrying about what would make the best evaluator. Come up with a simple one at first, and if you have
time you might think about how it could be improved. Just make sure that it bears some relation to the rules of the
game, and, most importantly, that it evaluates to a very large positive number (outside your normal range) when you
reach a winning position, and to a very large negative number for a losing position.<BR><BR>You must decide how
deep your program will search, but this decision should be encapsulated in a global variable that can easily be changed.
You should set it to search at least 3-ply deep, but feel free to set it higher if the program speed is reasonable.</TT>
]]></tei:div>
                                                <tei:div type="details">
                                                        <gsd_cw:attach
ref="www.people.fas.harvard.edu/~albert/cscie220/asst3.html"/>
                                                </tei:div>
                                                <!-- projects & grades -->
                                                <tei:note type="assigned">
                                                        <tei:date value="2002-10-22 12:00
AM">2002-10-22 12:00 AM</tei:date>
                                                </tei:note>
                                                <tei:note type="due">
                                                        <tei:date value="2002-10-22 9:00
AM">2002-10-22 09:00 AM</tei:date>
                                                </tei:note>
                                        </tei:cell>
                                        <tei:cell>
                                                <tei:head>Short Exercise in Parsing</tei:head>
                                                <tei:div type="description"><![CDATA[
<TT>Modify the provided grammar so that the program parse a richer subset of the English language and permits
multiple parses for at least one sentence (i.e. it allows for ambiguities). You must add production rules, new
preterminals, and additional lexical items (that is, English words). Try to maintain proper English grammar, so that
non-syntactic sentences will not parse. You will not need to make any modifications to the program other than to the
setf property-list definitions at the top of the file, and to the TERMINALP function.<BR><BR>Add at least four rules,
two preterminals, and six words to the language; feel free to add more as desired. Ambiguities are easiest to create from
prepositional phrases and from conjunctions.</TT>
]]></tei:div>
                                                <!-- projects & grades -->
                                                <tei:note type="assigned">
                                                        <tei:date value="2002-10-22 12:00
AM">2002-10-22 12:00 AM</tei:date>
                                                </tei:note>
                                                <tei:note type="due">
                                                        <tei:date value="2002-10-22 2:00
PM">2002-10-22 02:00 PM</tei:date>
                                                </tei:note>
                                        </tei:cell>
                                </tei:row>
```

```
                                        <tei:row role="data">
                                                <tei:cell>
                                                        <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">one itg</tei:name>
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
                                                        <tei:list>
                                                                <tei:head>Grades</tei:head>
                                                                <tei:label>score</tei:label>
                                                                <tei:item>3</tei:item>
                                                        </tei:list>
                                                        <!-- project links -->
        No submission required
        <!-- comments -->
                                                        <tei:div type="comments"><![CDATA[ tough
break. make sure you do the readings before each class. ]]></tei:div>
                                                        <!-- submission info -->
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
        Grade not available
       <!-- project links -->
        Not submitted
        <!-- comments -->
                                                        <!-- submission info -->
                                                        <tei:note>Not Submitted</tei:note>
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
        Grade not available
       <!-- project links -->
        Not submitted
        <!-- comments -->
                                                        <!-- submission info -->
                                                        <tei:note>Not Submitted</tei:note>
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
        Grade not available
       <!-- project links -->
                                                        <gsd_cw:attach ref="http://www.harvard.edu"/>
                                                        <!-- comments -->
                                                        <!-- submission info -->
                                                </tei:cell>
                                        </tei:row>
                                        <tei:row role="data">
                                                <tei:cell>
                                                        <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">student itg</tei:name>
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
                                                        <tei:list>
                                                                <tei:head>Grades</tei:head>
                                                                <tei:label>score</tei:label>
                                                                <tei:item>9</tei:item>
                                                        </tei:list>
                                                        <!-- project links -->
        No submission required
        <!-- comments -->
```

```
                <tei:div type="comments"><![CDATA[ nice
work. ]]></tei:div>
                <!-- submission info -->
            </tei:cell>
            <tei:cell>
                <!-- grades -->
    Grade not available
   <!-- project links -->
    Not submitted
    <!-- comments -->
                <!-- submission info -->
                <tei:note>Not Submitted</tei:note>
            </tei:cell>
            <tei:cell>
                <!-- grades -->
    Grade not available
   <!-- project links -->
    Not submitted
    <!-- comments -->
                <!-- submission info -->
                <tei:note>Not Submitted</tei:note>
            </tei:cell>
            <tei:cell>
                <!-- grades -->
    Grade not available
   <!-- project links -->
    Not submitted
    <!-- comments -->
                <!-- submission info -->
                <tei:note>Not Submitted</tei:note>
            </tei:cell>
        </tei:row>
        <tei:row role="data">
            <tei:cell>
                <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">three itg</tei:name>
            </tei:cell>
            <tei:cell>
                <!-- grades -->
                <tei:list>
                    <tei:head>Grades</tei:head>
                    <tei:label>score</tei:label>
                    <tei:item>9</tei:item>
                </tei:list>
                <!-- project links -->
    No submission required
    <!-- comments -->
                <tei:div type="comments"><![CDATA[ nice
work. ]]></tei:div>
                <!-- submission info -->
            </tei:cell>
            <tei:cell>
                <!-- grades -->
                <tei:list>
                    <tei:head>Grades</tei:head>
                    <tei:label>function</tei:label>
                    <tei:item>7</tei:item>
                    <tei:label>style</tei:label>
                    <tei:item>8</tei:item>
                </tei:list>
                <!-- project links -->
```

```
                                                <gsd_cw:attach
ref="/2002/fall/101/cw/cw_assignments/F8BF72E3-E125-11D6-976900D0B7741CA9/"/>
                                                <!-- comments -->
                                                <tei:div type="comments"><![CDATA[ solid
work. remember, lisp is not C.  ]]></tei:div>
                                                <!-- submission info -->
                                        </tei:cell>
                                        <tei:cell>
                                                <!-- grades -->
                                                <tei:list>
                                                        <tei:head>Grades</tei:head>
                                                        <tei:label>function</tei:label>
                                                        <tei:item>5</tei:item>
                                                        <tei:label>style</tei:label>
                                                        <tei:item>4</tei:item>
                                                </tei:list>
                                                <!-- project links -->
                                                <gsd_cw:attach
ref="http://www.gsd.harvard.edu/~Student_Name/search/bfs.l"/>
                                                <!-- comments -->
                                                <tei:div type="comments"><![CDATA[ i think
you didn't read the assignment thoroughly. please try again.  ]]></tei:div>
                                                <!-- submission info -->
                                        </tei:cell>
                                        <tei:cell>
                                                <!-- grades -->
        Grade not available
       <!-- project links -->
        Not submitted
        <!-- comments -->
                                                <!-- submission info -->
                                                <tei:note>Not Submitted</tei:note>
                                        </tei:cell>
                                </tei:row>
                                <tei:row role="data">
                                        <tei:cell>
                                                <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">two itg</tei:name>
                                        </tei:cell>
                                        <tei:cell>
                                                <!-- grades -->
                                                <tei:list>
                                                        <tei:head>Grades</tei:head>
                                                        <tei:label>score</tei:label>
                                                        <tei:item>7</tei:item>
                                                </tei:list>
                                                <!-- project links -->
        No submission required
        <!-- comments -->
                                                <tei:div type="comments"><![CDATA[ make
sure you understand the readings. if you have questions, try to attend TA hours. ]]></tei:div>
                                                <!-- submission info -->
                                        </tei:cell>
                                        <tei:cell>
                                                <!-- grades -->
                                                <tei:list>
                                                        <tei:head>Grades</tei:head>
                                                        <tei:label>function</tei:label>
                                                        <tei:item>7</tei:item>
                                                        <tei:label>style</tei:label>
                                                        <tei:item>7</tei:item>
                                                </tei:list>
```

```
                                                        <!-- project links -->
                                                        <gsd_cw:attach
ref="/2002/fall/101/cw/cw_assignments/F8BF72F4-E125-11D6-976900D0B7741CA9/"/>
                                                        <!-- comments -->
                                                        <tei:div type="comments"><![CDATA[ nice
picture. ]]></tei:div>
                                                        <!-- submission info -->
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
                                                        <tei:list>
                                                                <tei:head>Grades</tei:head>
                                                                <tei:label>function</tei:label>
                                                                <tei:item>7</tei:item>
                                                                <tei:label>style</tei:label>
                                                                <tei:item>7</tei:item>
                                                        </tei:list>
                                                        <!-- project links -->
                                                        <gsd_cw:attach
ref="http://www.harvard.edu/404.html"/>
                                                        <!-- comments -->
                                                        <tei:div type="comments"><![CDATA[ watch
your error checking. try again? ]]></tei:div>
                                                        <!-- submission info -->
                                                        <tei:note>Re-Submitted</tei:note>
                                                </tei:cell>
                                                <tei:cell>
                                                        <!-- grades -->
          Grade not available
         <!-- project links -->
          Not submitted
          <!-- comments -->
                                                        <!-- submission info -->
                                                        <tei:note>Not Submitted</tei:note>
                                                </tei:cell>
                                        </tei:row>
                                </tei:table>
                        </tei:div>
                        <!-- collections -->
                        <tei:div id="collections">
                                <tei:div type="collection">
                                        <tei:head>sample collection</tei:head>
                                        <tei:div type="collection_photo">
                                                <tei:head>first picture (Studnet_Name)</tei:head>
                                        </tei:div>
                                        <tei:div type="collection_photo">
                                                <tei:head>photoshop test (Studnet_Name)</tei:head>
                                        </tei:div>
                                </tei:div>
                        </tei:div>
                        <!-- /collections -->
                        <tei:div id="mails">
                                <tei:div type="mail">
                                        <tei:head>zak tests (<tei:date value="2002-11-04 12:33 PM">2002-
11-04 12:33 PM</tei:date>
                                                <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Student_Name</tei:name>)</tei:head>
                                        <tei:list>
                                                <tei:head>Recipients</tei:head>
                                                <tei:item>student one
        <tei:list>
                                                                <tei:item>
```

```
                                                                    <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">one itg</tei:name>
                                                                </tei:item>
                                                            </tei:list>
                                                        </tei:item>
                                                    </tei:list><![CDATA[
<TT>here is a message</TT>
]]><tei:list>
                                                        <tei:head>Access Log</tei:head>
                                                    </tei:list>
                                                </tei:div>
                                                <tei:div type="mail">
                                                    <tei:head>zak tests attach (<tei:date value="2002-11-04 12:33
PM">2002-11-04 12:33 PM</tei:date>
                                                    <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Studnet_Name</tei:name>)</tei:head>
                                                    <tei:list>
                                                        <tei:head>Recipients</tei:head>
                                                        <tei:item>student one
        <tei:list>
                                                                <tei:item>
                                                                    <tei:name type="person"
key="Student_Name+itg1@gsd.harvard.edu" gsd_cw:huid="">one itg</tei:name>
                                                                </tei:item>
                                                            </tei:list>
                                                        </tei:item>
                                                    </tei:list><![CDATA[
<TT>here is a message with attachments</TT>
]]><tei:div type="attachments">
                                                        <tei:head>Attachments</tei:head>
                                                        <gsd_cw:attach
ref="/2002/fall/101/cw/cw_mail/245B3A20-EF9A-11D6-
976A00D0B7741CA9/internet_map.gif?224401976478328558">internet_map.gif <tei:date value="2002-11-04 1:46
PM">2002-11-04 01:46 PM</tei:date>
                                                        </gsd_cw:attach>
                                                        <gsd_cw:attach
ref="/2002/fall/101/cw/cw_mail/245B3A20-EF9A-11D6-
976A00D0B7741CA9/sam_i_am.jpg?224401976478328558">sam_i_am.jpg <tei:date value="2002-11-04 1:47
PM">2002-11-04 01:47 PM</tei:date>
                                                        </gsd_cw:attach>
                                                    </tei:div>
                                                    <tei:list>
                                                        <tei:head>Access Log</tei:head>
                                                    </tei:list>
                                                </tei:div>
                                                <tei:div type="mail">
                                                    <tei:head>zak tests (<tei:date value="2002-11-21 8:35 PM">2002-
11-21 08:35 PM</tei:date>
                                                    <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Studnet_Name</tei:name>)</tei:head>
                                                    <tei:list>
                                                        <tei:head>Recipients</tei:head>
                                                        <tei:item>student
        <tei:list>
                                                                <tei:item>
                                                                    <tei:name type="person"
key="Student_Name+itg1@gsd.harvard.edu" gsd_cw:huid="">one itg</tei:name>
                                                                </tei:item>
                                                                <tei:item>
                                                                    <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">student itg</tei:name>
                                                                </tei:item>
```

```
                                                                        <tei:item>
                                                                                <tei:name type="person"
key="Student_Name+itg3@gsd.harvard.edu" gsd_cw:huid="">three itg</tei:name>
                                                                        </tei:item>
                                                                        <tei:item>
                                                                                <tei:name type="person"
key="Student_Name+itg2@gsd.harvard.edu" gsd_cw:huid="">two itg</tei:name>
                                                                        </tei:item>
                                                                </tei:list>
                                                        </tei:item>
                                                </tei:list><![CDATA[
<TT>zak tests</TT>
]]><tei:list>
                                                        <tei:head>Access Log</tei:head>
                                                </tei:list>
                                        </tei:div>
                                        <tei:div type="mail">
                                                <tei:head>zak is testing (<tei:date value="2002-11-21 8:59
PM">2002-11-21 08:59 PM</tei:date>
                                                        <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Studnet_Name</tei:name>)</tei:head>
                                                <tei:list>
                                                        <tei:head>Recipients</tei:head>
                                                        <tei:item>student
        <tei:list>
                                                                        <tei:item>
                                                                                <tei:name type="person"
key="Student_Name+itg1@gsd.harvard.edu" gsd_cw:huid="">one itg</tei:name>
                                                                        </tei:item>
                                                                        <tei:item>
                                                                                <tei:name type="person"
key="Student_Name@gsd.harvard.edu" gsd_cw:huid="">student itg</tei:name>
                                                                        </tei:item>
                                                                        <tei:item>
                                                                                <tei:name type="person"
key="Student_Name+itg3@gsd.harvard.edu" gsd_cw:huid="">three itg</tei:name>
                                                                        </tei:item>
                                                                        <tei:item>
                                                                                <tei:name type="person"
key="Student_Name+itg2@gsd.harvard.edu" gsd_cw:huid="">two itg</tei:name>
                                                                        </tei:item>
                                                                </tei:list>
                                                        </tei:item>
                                                </tei:list><![CDATA[
<TT>zak is testing</TT>
]]><tei:list>
                                                        <tei:head>Access Log</tei:head>
                                                </tei:list>
                                        </tei:div>
                                </tei:div>
                                <!-- surveys -->
                                <tei:div id="surveys">
                                        <tei:head>Survey</tei:head>
                                        <tei:div type="survey">
                                                <tei:head>li&amp;apos;s survey</tei:head>
        here is my survey
        <tei:list>
                                                        <tei:head>Questions</tei:head>
                                                        <tei:item>
          multiple choise
          <tei:table rows="4" cols="5">
                                                                        <tei:row role="label">
```

<tei:cell>Respondent</tei:cell>
<tei:cell>red</tei:cell>
<tei:cell>green</tei:cell>
<tei:cell>blue</tei:cell>
<tei:cell>Other</tei:cell>
</tei:row>
<tei:row role="data">
<tei:cell>
<tei:name type="person" key="Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Studnet_Name</tei:name>
</tei:cell>
<tei:cell>1</tei:cell>
<tei:cell>1</tei:cell>
<tei:cell/>
<tei:cell/>
</tei:row>
<tei:row role="data">
<tei:cell>
<tei:name type="person" key="Student_Name@gsd.harvard.edu" gsd_cw:huid="HUID_Number">Studnet_Name</tei:name>
</tei:cell>
<tei:cell>1</tei:cell>
<tei:cell/>
<tei:cell/>
<tei:cell/>
</tei:row>
<tei:row role="data">
<tei:cell>
<tei:name type="person" key="Student_Name+itg1@gsd.harvard.edu" gsd_cw:huid="">one itg</tei:name>
</tei:cell>
<tei:cell>1</tei:cell>
<tei:cell/>
<tei:cell/>
<tei:cell/>
</tei:row>
</tei:table>
</tei:item>
<tei:item>
radio
<tei:table rows="2" cols="4">
<tei:row role="label">

<tei:cell>Respondent</tei:cell>
<tei:cell>first one</tei:cell>
<tei:cell>second one</tei:cell>
<tei:cell>Other</tei:cell>
</tei:row>
<tei:row role="data">
<tei:cell>
<tei:name type="person" key="Student_Name+itg3@gsd.harvard.edu" gsd_cw:huid="">three itg</tei:name>
</tei:cell>
<tei:cell/>
<tei:cell>1</tei:cell>
<tei:cell/>
</tei:row>
</tei:table>
</tei:item>
<tei:item>

```
        essay question
        <tei:list>
                                                                                <tei:label>
                                                                                        <tei:name type="person"
key="Student_Name+itg3@gsd.harvard.edu" gsd_cw:huid="">three itg</tei:name>
                                                                                </tei:label>
                                                                                <tei:item>my essy response sucks. my
essy response sucks. my essy response sucks. my essy response sucks. </tei:item>
                                                                                        </tei:list>
                                                                </tei:item>
                                                        </tei:list>
                                                </tei:div>
                                        </tei:div>
                                        <!-- /surveys -->
                                        <!-- threads -->
                                        <!-- /threads -->
                                        <!-- /TOC items -->
                                        <!-- announcements -->
                                        <!-- /announcements -->
                                </tei:body>
                                <tei:back>
   </tei:back>
                </tei:text>
</tei:TEI.2>
```

Example: XSL Stylesheet for GSD Test Course

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet
 version="1.0"
 xmlns:xsl = "http://www.w3.org/1999/XSL/Transform"
 xmlns:tei="http://tei-c.org.uk/Lite/"
 xmlns:gsd_cw="http://www.gsd.harvard.edu/library/information_systems/courseware_archive.html">
 <xsl:output method="xml" indent="yes" />
 <xsl:strip-space elements="*"/>


 <xsl:template match="/">
  <xsl:apply-templates />
 </xsl:template>


 <xsl:template match="tei:TEI.2">
  <course>
   <xsl:apply-templates />
  </course>
 </xsl:template>


 <xsl:template match="tei:teiHeader">
  <head>
  <xsl:apply-templates/>
  </head>
 </xsl:template>


 <xsl:template match="tei:fileDesc">
  <xsl:apply-templates/>
 </xsl:template>


 <xsl:template match="tei:title">
  <title><xsl:value-of select="." /></title>
 </xsl:template>


 <xsl:template match="tei:author">
  <author><xsl:apply-templates/></author>
 </xsl:template>


 <xsl:template match="tei:sponsor">
  <sponsor><xsl:value-of select="." /></sponsor>
 </xsl:template>


 <xsl:template match="tei:publicationStmt">
 </xsl:template>


 <xsl:template match="tei:sourceDesc">
 </xsl:template>


 <xsl:template match="tei:text">
  <xsl:apply-templates/>
 </xsl:template>
```

```
 <xsl:template match="tei:front">
  <courseDescription>
    <xsl:value-of select="." />
  </courseDescription>
 </xsl:template>


 <xsl:template match="tei:body">
  <xsl:apply-templates/>
 </xsl:template>


 <xsl:template match="tei:div[@type='file']">
  <file>
   <xsl:apply-templates/>
  </file>
 </xsl:template>


 <xsl:template match="tei:div[@id='announcements']">
  <announcements>
   <xsl:apply-templates/>
  </announcements>
 </xsl:template>


 <xsl:template match="tei:div[@type='announcement']">
  <announcement>
      <xsl:attribute name="priority"><xsl:value-of select="tei:note[@type='priority']"/></xsl:attribute>
   <xsl:apply-templates/>
  </announcement>
 </xsl:template>


 <xsl:template match="tei:note[@type='priority']">
 </xsl:template>


 <xsl:template match="tei:div[@type='attachments']">
  <attachments>
    <xsl:apply-templates/>
  </attachments>
 </xsl:template>


 <xsl:template match="tei:div[@id='assignments']">
  <assignments>
    <xsl:apply-templates/>
  </assignments>
 </xsl:template>


 <xsl:template match="tei:div[@id='assignments']/tei:table/tei:row[@role='label']/tei:cell[position() &gt; 1]">
  <assignment>
   <!-- assigned -->
   <xsl:attribute name="assigned"><xsl:value-of select="./tei:note[@type='assigned']/tei:date/@value"
/></xsl:attribute>
   <!-- due -->
   <xsl:attribute name="due"><xsl:value-of select="./tei:note[@type='due']/tei:date/@value" /></xsl:attribute>
   <xsl:apply-templates select="tei:head" />
```

```
  <xsl:apply-templates select="tei:div" />
  <!-- grades -->
  <received_work>
   <xsl:apply-templates select="../../tei:row[@role='data']" mode="received_work">
     <xsl:with-param name="assignment" select="position()"></xsl:with-param>
   </xsl:apply-templates>
  </received_work>
 </assignment>
</xsl:template>


<xsl:template match="tei:div[@id='assignments']/tei:table/tei:row[@role='data']" mode="received_work">
  <xsl:param name="assignment"/>
  <received>
   <student>
    <xsl:apply-templates select="tei:cell[1]" />
   </student>
   <xsl:apply-templates select="tei:cell[$assignment]" mode="received_work" />
  </received>
</xsl:template>


<xsl:template match="tei:cell" mode="received_work">
  <content>
   <!-- submission info -->
   <xsl:choose>
    <xsl:when test="contains(child::text(), 'No submission required')"/>
    <xsl:when test="contains(tei:note, 'Late')">
      <xsl:attribute name="submitted">Late</xsl:attribute>
    </xsl:when>
    <xsl:when test="contains(tei:note, 'Not Submitted')">
      <xsl:attribute name="submitted">Not Submitted</xsl:attribute>
    </xsl:when>
    <xsl:when test="contains(tei:note, 'Re-Submitted')">
      <xsl:attribute name="submitted">Re-Submitted</xsl:attribute>
    </xsl:when>
    <xsl:otherwise>
      <xsl:attribute name="submitted">On Time</xsl:attribute>
    </xsl:otherwise>
   </xsl:choose>
   <!-- work -->
   <xsl:choose>
    <xsl:when test="count(gsd_cw:attach)=1">
       <xsl:apply-templates select="gsd_cw:attach"/>
    </xsl:when>
    <xsl:when test="contains(child::text(), 'No submission required')">
     <xsl:attribute name="non_required">true</xsl:attribute>
    </xsl:when>
    <xsl:otherwise />
   </xsl:choose>
  </content>
  <!-- grades -->
  <xsl:apply-templates select="tei:list[contains(tei:head, 'Grades')]" mode="received_work" />
  <!-- comments -->
  <xsl:apply-templates select="tei:div"/>
</xsl:template>


<xsl:template match="tei:list" mode="received_work">
  <grades>
   <xsl:apply-templates select="tei:label" mode="received_work"/>
  </grades>
```

```
</xsl:template>

<xsl:template match="tei:label" mode="received_work">
  <grade name="{child::text()}"><xsl:value-of select="following-sibling::tei:item[1]"/></grade>
</xsl:template>

<xsl:template match="tei:div[@id='assignments']/tei:table/tei:row[@role='label']/tei:cell[position()=1]" />

<xsl:template match="tei:div[@id='assignments']/tei:table/tei:row[@role='data']" />

<xsl:template match="tei:div[@type='description']">
  <description>
   <xsl:apply-templates/>
  </description>
</xsl:template>

<xsl:template match="tei:div[@type='details']">
  <details>
   <xsl:apply-templates/>
  </details>
</xsl:template>

<xsl:template match="tei:div[@type='comments']">
  <comments>
   <xsl:apply-templates/>
  </comments>
</xsl:template>

<xsl:template match="tei:div[@id='mails']">
  <mails>
   <xsl:apply-templates/>
  </mails>
</xsl:template>

<xsl:template match="tei:div[@type='mail']">
  <mail>
   <xsl:apply-templates select="tei:list[tei:head/text() = 'Recipients']"/>
   <message><xsl:value-of select="child::text()"/></message>
   <xsl:apply-templates select="tei:list[tei:head/text() = 'Access Log']"/>
   <xsl:apply-templates select="tei:div"/>
  </mail>
</xsl:template>

<xsl:template match="tei:div[@type='mail']/tei:list[tei:head/text() = 'Recipients']">
 <recipients>
  <xsl:for-each select="tei:item">
   <recipient-group>
    <xsl:attribute name="name"><xsl:value-of select="substring-before(child::text()[1],' &#xA;')"/></xsl:attribute>
    <xsl:apply-templates select="tei:list/tei:item"/>
   </recipient-group>
  </xsl:for-each>
 </recipients>
```

```
</xsl:template>


<xsl:template match="tei:div[@type='mail']/tei:list[tei:head/text() = 'Access Log']">
  <xsl:if test="count(.//tei:item) &gt; 0">
  <access_log>
   <xsl:for-each select="tei:item">
    <access>
     <xsl:attribute name="timestamp"><xsl:value-of select="tei:date[1]/@value" /></xsl:attribute>
     <xsl:apply-templates select="tei:name"/>
    </access>
   </xsl:for-each>
  </access_log>
 </xsl:if>
</xsl:template>


<xsl:template match="tei:div[@id='surveys']">
 <surveys>
  <xsl:apply-templates />
 </surveys>
</xsl:template>


<xsl:template match="tei:div[@type='survey']">
 <survey>
  <xsl:apply-templates select="tei:head"/>
  <description><xsl:value-of select="child::text()[1]"/></description>
  <questions>
    <xsl:for-each select="tei:list/tei:item">
      <question>
       <question_text><xsl:value-of select="child::text()[1]"/></question_text>
        <xsl:choose>
         <xsl:when test="count(tei:table) &gt; 0">
           <!-- multiple choice or checkbox -->
           <xsl:for-each select="tei:table/tei:row[@role='data']">
            <response>
             <respondent><xsl:apply-templates select="tei:cell[1]"/></respondent>
             <xsl:for-each select="tei:cell[position() &gt; 1]">
               <xsl:variable name="pos" select="position() + 1" />
               <xsl:if test="child::text()='1'">
                <answer><xsl:value-of select="../../tei:row[1]/tei:cell[$pos]" /></answer>


               </xsl:if>
             </xsl:for-each>
            </response>
           </xsl:for-each>
         </xsl:when>
         <xsl:when test="count(tei:list) &gt ; 0">
           <!-- text -->
           <xsl:for-each select="tei:list/tei:label">
             <response>
               <respondent><xsl:apply-templates select="."/></respondent>
               <answer><xsl:value-of select="following-sibling: :tei:item[1]" /></answer>
             </response>
           </xsl:for-each>
         </xsl:when>
        </xsl:choose>
      </question>
    </xsl:for-each>
  </questions>
```

```
 </survey>
</xsl:template>


<xsl:template match="tei:div[@id='collections']">
 <collections>
  <xsl:apply-templates />
 </collections>
</xsl:template>


<xsl:template match="tei:div[@type='collection']">
 <collection>
  <xsl:apply-templates />
 </collection>
</xsl:template>


<xsl:template match="tei:div[@type='collection_photo']">
 <collection_photo>
  <xsl:apply-templates />
 </collection_photo>
</xsl:template>


<xsl:template match="tei:div">
 <undifferentiated_div />
</xsl:template>


<!-- head: significant processing needed -->
<xsl:template match="tei:head">
 <xsl:choose>
  <xsl:when test="count(tei:date) = 2">
   <!-- begin and end date -->
   <xsl:attribute name="start_date"><xsl:value-of select="tei:date[1]/@value"/></xsl:attribute>
   <xsl:attribute name="end_date"><xsl:value-of select="tei:date[2]/@value"/></xsl:attribute>
   <subtitle><xsl:value-of select="substring(./child::text()[1], 1, string-length(./child::text()[1])-2)"/></subtitle>
   <author>
    <xsl:apply-templates select="tei:name"/>
   </author>
  </xsl:when>
  <xsl:when test="count(tei:date) = 1">
   <!-- single date -->
   <xsl:attribute name="date"><xsl:value-of select="tei:date[1]/@value"/></xsl:attribute>
   <subtitle><xsl:value-of select="substring(./child::text()[1], 1, string-length(./child::text()[1])-2)"/></subtitle>
   <author>
    <xsl:apply-templates select="tei:name"/>
   </author>
  </xsl:when>
  <xsl:otherwise>
   <subtitle><xsl:value-of select="." /></subtitle>
  </xsl:otherwise>
 </xsl:choose>
</xsl:template>


<xsl:template match="gsd_cw:attach">
 <attachment location="{@ref}">
  <xsl:if test="@alt">
   <xsl:attribute name="alternate"><xsl:value-of select="@alt"/></xsl:attribute>
  </xsl:if>
```

```
    <!-- contents -->
    <xsl:value-of select="." />
    </attachment>
  </xsl:template>


  <xsl:template match="tei:name">
    <person email="{@key}" huid="{@gsd_cw:huid}"><xsl:value-of select="." /></person>
  </xsl:template>


</xsl:stylesheet>
```

Example: Result of XML file with Applied XSL for GSD Test Course


```
<?xml version="1.0" encoding="utf-8"?>
<course xmlns:tei="http://tei-c.org.uk/Lite/"
xmlns:gsd_cw="http://www.gsd.harvard.edu/library/information_systems/courseware_archive.html">
	<head>
		<title>101:Test Course: Independent Study Fall 2002:machine-readable form of instructional
website</title>
		<author>
			<person email="cwadmin@gsd.harvard.edu" huid="">instructor itg</person>
		</author>
		<sponsor>Harvard Graduate School of Design</sponsor>
	</head>
	<courseDescription/>
	<file>
		<subtitle>Syllabus</subtitle>
		<attachment location="2002/fall/101/cw/syllabus.c"/>
	</file>
	<announcements>
		<announcement priority="4" start_date="2003-01-10 4:00 PM" end_date="2003-01-31 4:00 PM">
			<subtitle>testing jan 31</subtitle>
			<author>
				<person email="Student_Name@gsd.harvard.edu"
huid="HUID_Number">Studnet_Name</person>
			</author>


&amp;lt;TT&amp;gt;testing&amp;lt;/TT&amp;gt;


	</announcement>
		<announcement priority="4" start_date="2002-11-01 4:00 PM" end_date="2003-11-01 4:00 PM">
			<subtitle>are annc's fixed?</subtitle>
			<author>
				<person email="Student_Name@gsd.harvard.edu"
huid="HUID_Number">Studnet_Name</person>
			</author>


&amp;lt;TT&amp;gt;this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are
annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test;
are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? &amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;this is a test; are
annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test;
are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed?
v&amp;lt;/TT&amp;gt;


	</announcement>
		<announcement priority="4" start_date="2002-11-01 4:00 PM" end_date="2003-11-01 4:00 PM">
			<subtitle>are annc's with attch's fixed?</subtitle>
			<author>
				<person email="Student_Name@gsd.harvard.edu"
huid="HUID_Number">Studnet_Name</person>
			</author>


&amp;lt;TT&amp;gt;this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are
annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test;
are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? &amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;this is a test; are
annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test;
are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed? this is a test; are annc's fixed?
v&amp;lt;/TT&amp;gt;
```

```
    <attachments>
                                    <subtitle>Attachments</subtitle>
                                    <attachment location="/2002/fall/101/cw/cw_announcements/23F044DC-
EC22-11D6-976A00D0B7741CA9/internet_map.gif?033701666748065041">internet_map.gif</attachment>
                                </attachments>
                        </announcement>
                        <announcement priority="4" start_date="2002-09-20 4:00 PM" end_date="2002-09-21 4:00 PM">
                                <subtitle>Test Announcement - no class today, 9/20th</subtitle>
                                <author>
                                        <person email="cwadmin@gsd.harvard.edu" huid="">instructor itg</person>
                                </author>


&amp;lt;TT&amp;gt;&amp;lt;/TT&amp;gt;


      </announcement>
          </announcements>
          <assignments>
                        <assignment assigned="2002-09-13 12:00 AM" due="2002-09-15 10:00 AM">
                                <subtitle>pop quiz</subtitle>
                                <description>


&amp;lt;TT&amp;gt;here are the results of the pop quiz from September 15. &amp;lt;/TT&amp;gt;


          </description>
                                <received_work>
                                        <received>
                                                <student>
                                                        <person email="Student_Name+itg1@gsd.harvard.edu"
huid="">one itg</person>
                                                </student>
                                                <content non_required="true"/>
                                                <grades>
                                                        <grade name="score">3</grade>
                                                </grades>
                                                <comments>
           tough break. make sure you do the readings before each class.
          </comments>
                                        </received>
                                        <received>
                                                <student>
                                                        <person email="Student_Name@gsd.harvard.edu"
huid="">student itg</person>
                                                </student>
                                                <content non_required="true"/>
                                                <grades>
                                                        <grade name="score">9</grade>
                                                </grades>
                                                <comments>
           nice work.
          </comments>
                                        </received>
                                        <received>
                                                <student>
                                                        <person email="Student_Name+itg3@gsd.harvard.edu"
huid="">three itg</person>
                                                </student>
```

```
                                        <content non_required="true"/>
                                        <grades>
                                                <grade name="score">9</grade>
                                        </grades>
                                        <comments>
        nice work.
        </comments>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg2@gsd.harvard.edu"
huid="">two itg</person>
                                        </student>
                                        <content non_required="true"/>
                                        <grades>
                                                <grade name="score">7</grade>
                                        </grades>
                                        <comments>
        make sure you understand the readings. if you have questions, try to attend TA hours.
        </comments>
                                </received>
                        </received_work>
                </assignment>
                <assignment assigned="2002-10-22 12:00 AM" due="2002-10-22 8:00 AM">
                        <subtitle>Graph Searching</subtitle>
                        <description>


&amp;lt;TT&amp;gt;You must write a total of four separate search programs: BREADTH-FIRST, DEPTH-FIRST, LEAST-
COST, and ASTAR. However, if you plan your code well, you may find that the four are nearly identical in most
respects, and in fact you may be able to write a single basic search function with four top-level function calls to
it.&amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;For your A* algorithm, use the air distances between two cities as the heuristic estimate for
the remaining cost.&amp;lt;/TT&amp;gt;


        </description>
                        <details>
                                <attachment
location="http://www.people.fas.harvard.edu/~albert/cscie220/asst2.html"/>
                        </details>
                        <received_work>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg1@gsd.harvard.edu"
huid="">one itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name@gsd.harvard.edu"
huid="">student itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg3@gsd.harvard.edu"
huid="">three itg</person>
                                        </student>
                                        <content submitted="On Time">
```

```
                                                <attachment
location="/2002/fall/101/cw/cw_assignments/F8BF72E3-E125-11D6-976900D0B7741CA9/"/>
                                        </content>
                                        <grades>
                                                <grade name="function">7</grade>
                                                <grade name="style">8</grade>
                                        </grades>
                                        <comments>
        solid work. remember, lisp is not C.
        </comments>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg2@gsd.harvard.edu"
huid="">two itg</person>
                                        </student>
                                        <content submitted="On Time">
                                                <attachment
location="/2002/fall/101/cw/cw_assignments/F8BF72F4-E125-11D6-976900D0B7741CA9/"/>
                                        </content>
                                        <grades>
                                                <grade name="function">7</grade>
                                                <grade name="style">7</grade>
                                        </grades>
                                        <comments>
        nice picture.
        </comments>
                                </received>
                        </received_work>
                </assignment>
                <assignment assigned="2002-10-22 12:00 AM" due="2002-10-22 9:00 AM">
                        <subtitle>Mimimax Searching</subtitle>
                        <description>


&amp;lt;TT&amp;gt;GET-MOVE and the associated functions you write must be an implementation of the minimax routine (alpha-beta pruning is an optional extension). You will need to write a set of functions to properly implement this routine, but GET-MOVE must be the top-level function and must work as stated above. If you take advantage of LISP's recursive nature you will find that this problem is very simple to program. A pair of mutually recursive functions can perform almost all of the work of maintaining the search tree for you, with no actual tree data structure required.&amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;You will need to write a function which "expands" a node, but notice that this task is fairly simple since all you need to do is check which moves are legal and then run %DO-MOVE on each of them, giving it the appropriate player (which will be different at each level of the tree), the hole number of the legal move, and the board position of the node you are expanding. (Don't send %DO-MOVE an illegal move! It won't check, and anything might happen!)&amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;Since minimax relies on a heuristic evaluator to avoid searching all the way to the end of the game, you will need to write such an evaluator. Unless you find yourself enchanted by the prospect, you needn't spend too much time worrying about what would make the best evaluator. Come up with a simple one at first, and if you have time you might think about how it could be improved. Just make sure that it bears some relation to the rules of the game, and, most importantly, that it evaluates to a very large positive number (outside your normal range) when you reach a winning position, and to a very large negative number for a losing position.&amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;You must decide how deep your program will search, but this decision should be encapsulated in a global variable that can easily be changed. You should set it to search at least 3-ply deep, but feel free to set it higher if the program speed is reasonable.&amp;lt;/TT&amp;gt;


        </description>
                        <details>
                                <attachment
location="www.people.fas.harvard.edu/~albert/cscie220/asst3.html"/>
                        </details>
                        <received_work>
                                <received>
```

```
                                        <student>
                                                <person email="Student_Name+itg1@gsd.harvard.edu"
huid="">one itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name@gsd.harvard.edu"
huid="">student itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg3@gsd.harvard.edu"
huid="">three itg</person>
                                        </student>
                                        <content submitted="On Time">
                                                <attachment
location="http://www.gsd.harvard.edu/~Student_Name/search/bfs.l"/>
                                        </content>
                                        <grades>
                                                <grade name="function">5</grade>
                                                <grade name="style">4</grade>
                                        </grades>
                                        <comments>
        i think you didn't read the assignment thoroughly. please try again.
        </comments>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg2@gsd.harvard.edu"
huid="">two itg</person>
                                        </student>
                                        <content submitted="Re-Submitted">
                                                <attachment
location="http://www.harvard.edu/404.html"/>
                                        </content>
                                        <grades>
                                                <grade name="function">7</grade>
                                                <grade name="style">7</grade>
                                        </grades>
                                        <comments>
        watch your error checking. try again?
        </comments>
                                </received>
                        </received_work>
                </assignment>
                <assignment assigned="2002-10-22 12:00 AM" due="2002-10-22 2:00 PM">
                        <subtitle>Short Exercise in Parsing</subtitle>
                        <description>


&amp;lt;TT&amp;gt;Modify the provided grammar so that the program parse a richer subset of the English language and permits multiple parses for at least one sentence (i.e. it allows for ambiguities). You must add production rules, new preterminals, and additional lexical items (that is, English words). Try to maintain proper English grammar, so that non-syntactic sentences will not parse. You will not need to make any modifications to the program other than to the setf property-list definitions at the top of the file, and to the TERMINALP function.&amp;lt;BR&amp;gt;&amp;lt;BR&amp;gt;Add at least four rules, two preterminals, and six words to the language; feel free to add more as desired. Ambiguities are easiest to create from prepositional phrases and from conjunctions.&amp;lt;/TT&amp;gt;
```

```
        </description>
                        <received_work>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg1@gsd.harvard.edu"
huid="">one itg</person>
                                        </student>
                                        <content submitted="On Time">
                                                <attachment location="http://www.harvard.edu"/>
                                        </content>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name@gsd.harvard.edu"
huid="">student itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg3@gsd.harvard.edu"
huid="">three itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                                <received>
                                        <student>
                                                <person email="Student_Name+itg2@gsd.harvard.edu"
huid="">two itg</person>
                                        </student>
                                        <content submitted="Not Submitted"/>
                                </received>
                        </received_work>
                </assignment>
        </assignments>
        <collections>
                <collection>
                        <subtitle>sample collection</subtitle>
                        <collection_photo>
                                <subtitle>first picture (Studnet_Name)</subtitle>
                        </collection_photo>
                        <collection_photo>
                                <subtitle>photoshop test (Studnet_Name)</subtitle>
                        </collection_photo>
                </collection>
        </collections>
        <mails>
                <mail>
                        <recipients>
                                <recipient-group name="student one">
                                        <person email="Student_Name+itg1@gsd.harvard.edu" huid="">one
itg</person>
                                </recipient-group>
                        </recipients>
                        <message>


&amp;lt;TT&amp;gt;here is a message&amp;lt;/TT&amp;gt;


        </message>
```

```
		</mail>
		<mail>
			<recipients>
				<recipient-group name="student one">
					<person email="Student_Name+itg1@gsd.harvard.edu" huid="">one
itg</person>
				</recipient-group>
			</recipients>
			<message>


&amp;lt;TT&amp;gt;here is a message with attachments&amp;lt;/TT&amp;gt;


	</message>
			<attachments>
				<subtitle>Attachments</subtitle>
				<attachment location="/2002/fall/101/cw/cw_mail/245B3A20-EF9A-11D6-
976A00D0B7741CA9/internet_map.gif?224401976478328558">internet_map.gif 2002-11-04 01:46 PM</attachment>
				<attachment location="/2002/fall/101/cw/cw_mail/245B3A20-EF9A-11D6-
976A00D0B7741CA9/sam_i_am.jpg?224401976478328558">sam_i_am.jpg 2002-11-04 01:47 PM</attachment>
			</attachments>
		</mail>
		<mail>
			<recipients>
				<recipient-group name="student">
					<person email="Student_Name+itg1@gsd.harvard.edu" huid="">one
itg</person>
					<person email="Student_Name@gsd.harvard.edu" huid="">student
itg</person>
					<person email="Student_Name+itg3@gsd.harvard.edu"
huid="">three itg</person>
					<person email="Student_Name+itg2@gsd.harvard.edu" huid="">two
itg</person>
				</recipient-group>
			</recipients>
			<message>


&amp;lt;TT&amp;gt;zak tests&amp;lt;/TT&amp;gt;


	</message>
		</mail>
		<mail>
			<recipients>
				<recipient-group name="student">
					<person email="Student_Name+itg1@gsd.harvard.edu" huid="">one
itg</person>
					<person email="Student_Name@gsd.harvard.edu" huid="">student
itg</person>
					<person email="Student_Name+itg3@gsd.harvard.edu"
huid="">three itg</person>
					<person email="Student_Name+itg2@gsd.harvard.edu" huid="">two
itg</person>
				</recipient-group>
			</recipients>
			<message>


&amp;lt;TT&amp;gt;zak is testing&amp;lt;/TT&amp;gt;
```

```
	</message>
		</mail>
	</mails>
	<surveys>
		<subtitle>Survey</subtitle>
		<survey>
			<subtitle>li's survey</subtitle>
			<description>
	here is my survey
	</description>
			<questions>
				<question>
					<question_text>
	  multiple choise
	  </question_text>
					<response>
						<respondent>
							<person
email="Student_Name@gsd.harvard.edu" huid="HUID_Number">Studnet_Name</person>
						</respondent>
						<answer>red</answer>
						<answer>green</answer>
					</response>
					<response>
						<respondent>
							<person
email="Student_Name@gsd.harvard.edu" huid="HUID_Number">Studnet_Name</person>
						</respondent>
						<answer>red</answer>
					</response>
					<response>
						<respondent>
							<person
email="Student_Name+itg1@gsd.harvard.edu" huid="">one itg</person>
						</respondent>
						<answer>red</answer>
					</response>
				</question>
				<question>
					<question_text>
	  radio
	  </question_text>
					<response>
						<respondent>
							<person
email="Student_Name+itg3@gsd.harvard.edu" huid="">three itg</person>
						</respondent>
						<answer>second one</answer>
					</response>
				</question>
				<question>
					<question_text>
	  essay question
	  </question_text>
					<response>
						<respondent>
							<person
email="Student_Name+itg3@gsd.harvard.edu" huid="">three itg</person>
						</respondent>
						<answer>my essy response sucks. my essy response sucks.
my essy response sucks. my essy response sucks. </answer>
```

```
                                                    </response>
                                                </question>
                                            </questions>
                                        </survey>
                                    </surveys>
</course>
```

Report to the University

(Submitted to the LDI Steering Committee on December 12, 2003)

Request for Further Study to Address the Archiving of Selected Web-based Course Materials: Findings based on a Graduate School of Design Library Digital Initiative project

Report submitted on behalf of:

Graduate School of Design, Frances Loeb Library
Harvard Business School, Baker Library
Harvard College Library, Judaica Division
Harvard Law School Library
Harvard Medical School, Countway Library of Medicine
Harvard University Archives
Harvard University Library, Office for Information Systems
Harvard University Library, Weissman Preservation Center
iCommons

Table of Contents

Summary

This report is the result of a one-year project funded by the Harvard University Library's Library Digital Initiative (LDI) to examine the challenges posed by the preservation of web-based curriculum materials at Harvard.

Two project committees were formed to address policy and technical issues respectively, and used course web sites created at the Graduate School of Design (GSD) from 1999 to Spring 2003 as sample data. Project participants concluded that web-based course curricula represent a significant investment of University resources and now form an integral part of the educational fabric of the University. It is in the University's interest to address the long-term preservation of these materials.

> *As a result of this project, we strongly recommend that the University collect, preserve, and maintain access to web-based course curricula in a systematic manner.*

Such preservation should support the availability of these assets for ongoing use in teaching *and* to provide documentation of the material aspects of Harvard's teaching mission for the historical record and long-term research use.

In support of this goal we recommend:

- *Examination of the University's rights to preserve web-based course material and make them available for educational use and retain the non-exclusive right to copy or reformat the materials for the purposes of preservation and access.*
- *Establishment of University policies to address the confidentiality of student information and intellectual property rights of materials in course web sites.*
- *Development of appraisal guidelines by University archivists for the selection of web-based course curricula.*
- *Development of archiving functions, consistent with the technical preservation and access standards, for each of the course platforms deployed at Harvard.*

In support of these recommendations we urge the establishment of two University committees:

- *One committee to address the legal and administrative challenges of preserving web-based course materials. These include questions of responsibility for maintaining the materials, intellectual property rights, and confidentiality of student information.*
- *A second committee to address the technical challenges including the integration of archival functions into learning management systems deployed at Harvard.*

Participants of the project working groups:

Graduate School of Design, Frances Loeb Library
Harvard Business School, Baker Library
Harvard College Library, Judaica Division
Harvard Law School Library
Harvard Medical School, Countway Library of Medicine
Harvard University Archives
Harvard University Library, Office for Information Systems
Harvard University Library, Weissman Preservation Center
iCommons

Project background

In 2001 the Francis Loeb Library at the Graduate School of Design (GSD) received LDI funding for ***Web-based Course Material Archiving Project - Study Phase***. During this project the Loeb Library planned procedures and a technical infrastructure model for the documentation and archiving of digital course materials in various formats. Two advisory committees surveyed file formats and software applications used to create course web sites and investigated copyright issues, documentation standards, and use and access models. The project also involved the design of a framework for the technological infrastructure and implementation method to support the capture and preservation of course web sites.

Two categories of course materials were studied in this project:

3. Legacy course web sites dating from Fall 1999 through Spring 2003 that were saved on storage systems,
4. Current-use course materials created in a variety of platforms.

Two audiences were identified for archival course materials:

3. Faculty and students seeking to continue to use a particular course curriculum in its original or updated form,
4. Researchers studying the development of curricula, teaching methods, and academic disciplines at Harvard.

The technical advisory committee worked with the GSD's Head of Instructional Technology to formulate working models for preserving the legacy course materials and the current-use course materials. The functional advisory committee addressed the administrative and policy issues that affect the acquisition and preservation of web-based course materials.

As the result of this project, the advisory committees have made a number of recommendations in support of a University-wide model for collecting, preserving, and maintaining access to web-based course materials.

The importance of course materials

Academic endeavor, the teaching and research by faculty and students, distinguishes universities from other types of institutions and distinguishes a particular university from its peers. As its primary purpose an academic archive documents and preserves the identity and corporate memory of the institution, including the activities and decisions it undertakes in carrying out its academic mission: the education of students, the creation of new knowledge through research, and the sharing of that knowledge for the benefit of the institutional community and society at large. At the core of any university's corporate memory is the record of the courses taught in the university's degree programs.

Corporate memory serves a number of purposes. A clear view of the past provides a foundation and context for future decisions, chronicles the accomplishments of the institution and its constituents, and maintains the identity of the institution. It documents the role of the institution in the development of higher education and the progress of the professions. A university's historical record supports the long-term value of its graduates' credentials.

Researchers working in fields such as intellectual history, the history of higher education, and the rise of the professions have long appreciated the importance of syllabi, lecture notes, students' class notes, and other artifacts of teaching as a means of providing historical "snapshots" of the development of a discipline. The preservation of this type of material, however, has often been haphazard. Fortunate, indeed, are those educational institutions and professional associations that have systematically collected the evidence of what was being taught to its students and neophyte practitioners and the methods and successes of this pedagogy.

Course materials at Harvard

Harvard's faculty, administrators, and archivists have recognized the value of course materials, although the scope and completeness of course documentation varies with each school's academic culture, administrative resources, and opportunity for capture. Some of the course records are acquired by the University's archives but many remain active department files or are discarded once the immediate need has been met.

As a complicating factor, faculty members retain the intellectual property rights to some teaching materials. Such materials have not been defined as University records. Thus they have not always been the primary focus of archival collecting efforts.

Harvard's schools now emphasize and encourage the use of information technology to expand the classroom experience and facilitate communication among faculty and students. Course catalogs are available on the web and are fully searchable. Course web sites are increasingly common and several schools have established instructional technology programs to support the development and use of electronic teaching resources. Web-based curriculum materials may be shared, re-used and revised to form new course materials.

For example, at the Graduate School of Design faculty members often use web-based curriculum materials developed for a particular course as building blocks for new materials. The GSD's centrally-accessible portal for teaching resources has also encouraged the sharing, re-use, and further development of course materials.

With learning management systems now in place at the GSD, Business School, Medical School, and Faculty of Arts and Sciences, network-based teaching materials now represent a significant investment of the University's resources, from network upgrades and hardware purchases to labor on the part of IT specialists, teaching fellows, and faculty. Early experience shows, however, that these materials are some of the University's most evanescent assets, having been created in a wide variety of formats with storage and maintenance handled locally on a case-by-case basis.

Unlike traditional paper based materials, electronic resources will not stand the test of time if left unmanaged. The technical skills and costs needed to maintain these electronic resources may be prohibitive if each school is left to solve the problem on its own. The establishment of the Harvard Academic Computing Committee and the subsequent emergence of iCommons offer new opportunities to capture and preserve the academic experience in a systematic way.

> *We strongly recommend that the University take advantage of the opportunity provided by these campus-wide initiatives to collect, preserve, and maintain access to these pedagogic assets in a systematic manner.*

Challenges to be addressed

Three types of challenges dominate any attempt to preserve web-based teaching materials: legal, administrative, and technical. The University must address these challenges and develop long-term solutions before it embarks on a program to collect and preserve these materials.[4]

1. Legal challenges

These involve intellectual property rights to web-based teaching materials and privacy rights governing student information that may appear in course web sites. While archivists, technical specialists, and administrators can work to resolve the technical and logistical issues, policies addressing the legal issues

[4] For an excellent discussion of these issues, see *The Afterlives of Courses on the Nework: Information Management Issues for Learning Management Systems* by Clifford Lynch, EDUCAUSE Center for Applied Research Research Bulletin, Vol. 2002, Issue 23, November 26, 2002, http://www.cni.org/staff/cliffpubs/ECARpaper2002.pdf

must be established before the curriculum materials can be made available for use by the University community or outside researchers.

According to current University policy, members of the faculty retain the rights to their intellectual output. The iCommons courseware platform has been developed with the understanding that the University owns the courseware itself, while the faculties who use it own the content they have created. Before the University's archives devote any resources to acquiring web-based course materials, the right to preserve and maintain these materials must be clarified.

> *We recommend the examination of the University's interests and legal rights in web-based course materials to preserve selected teaching materials created at the University, and the establishment of the University's right to make them available for educational use, including the non-exclusive right to copy or reformat the materials for the purposes of preservation and access.*

All of the learning management system platforms in use at Harvard intersect with administrative and registrarial information systems at various schools. Much of the information provided by these systems is subject to the Family Educational Rights and Privacy Act (FERPA). Consideration must be given to the confidentiality of student information subject to FERPA that may be included in course web sites.

> *We recommend the establishment of University policies to protect the confidentiality of student information in course web sites selected for preservation.*

2. Administrative challenges

Administrative policies need to be established on the University and local level to guide the capture and preservation of web-based course materials before any implementation process can occur.

These policies must address the assignment of responsibility for web-based teaching materials both in current-use and long-term research environments. Some course web sites will be best preserved and served centrally by the appropriate archive. The school of origin must be responsible for the maintenance of course materials for ongoing teaching purposes.

Administrative challenges also involve the appraisal of web-based course materials in the context of short- and long-term University needs. The selection of course material for ongoing use and long-term preservation will impact the allocation of resources to implement the appropriate preservation and access methodologies, so guidelines to support sensible, sustainable decisions will be needed.

> *We recommend the University's archives create a set of appraisal guidelines to identify electronic web-based course materials with long-term significance to the University.*

3. Technical challenges

These involve the capture, storage, and preservation of course web sites, as well as procedures for making the materials accessible and functional in the future. Technical issues will vary, depending upon whether archivists and administrators are working with legacy teaching materials or current-use materials.

With the coordinated development and expanding implementation of the iCommons courseware platform and other learning management systems at the University, the establishment of systematic methods for long-term preservation and access is more likely to be achieved.

> *We recommend the development of technical standards for long-term preservation and access.*

> *We recommend the development of archiving functions, consistent with the technical standards for each of the courseware platforms deployed at Harvard.*

Recommendations

To implement these recommendations we urge the establishment of two University committees to develop and coordinate these policies and guidelines.

- One committee will be responsible for developing the recommended legal and administrative policies and guidelines.
- A second committee will be responsible for developing the recommended technical guidelines and specifications.

Graduate School of Design case study

Case study summary

As part of this project, technical specialists at the Graduate School of Design (GSD) surveyed materials in the Courseware system dating from Fall 1999 through Spring 2003, in order to assess the range of formats and possibilities for long-term preservation. One goal of the GSD was to document its own educational process. With an increase in electronic teaching tools it became apparent that a plan to capture and archive web-based course materials would need to be established in order to collect current materials. The purpose of the Library Digital Initiative (LDI) grant was to study the scope of the course materials and the options for capturing and archiving the online web-based content.

- In the course of this project, the GSD was able to create a model to assess and draft potential strategies to prepare the material created in Courseware for possible archival preservation. Flexible solutions were sought to ensure that future archival goals can be met. Full implementation of the model, however, will require a number of policy and case–specific decisions.

The GSD Courseware System

In 1999 the GSD developed Courseware, an in-house LMS that served as a single-entry portal to various pedagogical tools that provided GSD faculty with new ways to teach and deliver course materials. The Courseware system was designed with a back-end connection heavily dependent on the school's administrative database to gather student, course, and other administrative information to determine authentication, set levels of access, and integrate administrative data within course web sites. Courseware essentially worked as a Web-based learning system, delivering dynamically generated content data and access was determined by the GSD Registrar's database.

Occasionally users would upload files or post information back to the server through specific tools that further supplemented the content of the course such as discussion threads. Specific course content would occasionally be used as a building block for the course offered in a different term. The initial Courseware infrastructure lacked an adaptive learning environment for tracking learner progress and assessing content mastery, but had the potential with future development to include components that would allow for long-term storage.

Because of the demands posed by supporting the variety of formats used in many GSD courses and the school's goal of archiving this material, many difficult decisions had to be made that reflected the reality of the current technological and academic environment. The GSD tried to provide mechanisms that would provide flexibility and options for future archiving goals. Due to changes in the school's direction, development of Courseware was put on permanent hold, and eventually ended, during the Fall 2002 term as the school contemplated switching to the University's iCommons LMS.

Technical Solutions

Courseware was interwoven with the school's administrative database, and because of intellectual property rights issues, a process was needed to remove materials that were not actually part of a given course. The goal was to determine how to best utilize existing references provided by the administrative database, rather than examining all the individual files uploaded to the system and then determining the files' association with a respective course.

Given the limited resources available in the GSD, the project's technical specialists opted to base the preservation model on the use of services and expertise currently available to the Harvard community through Harvard University Library's Office of Information Systems (OIS) and on similar research and practice done outside the University. The preservation model involves use of the LDI Digital Repository Service (DRS) for long-term storage of digital files and an XML encoding schema for metadata. It also assumes that format migration will most likely be the method for maintaining the future usability of preserved materials.

A form of XML that would be readily usable, retain all course information, and could be customized in order to include tags specifically referencing associated files not housed in the database was the best

solution for storing the dynamically generated data for each class. One option would have been to create a project-specific DTD (document type definition) that mimics the structure of the Courseware records for an individual course. Another option was to use an open-source structure as a base. The base could be expanded while still allowing liberal use of type and identification attributes to distinguish various Courseware modules and their sub-elements, including a namespace to refer to attached files and an attribute to refer to users by their Harvard University ID numbers. Since the first approach requires a complex structure to be maintained indefinitely, the open-source option was chosen.

There were several available open-source metadata options including METS (Metadata Encoding & Transmission Standard), SCORM (Sharable Content Object Reference Model), and TEI-Lite (from the Text Encoding Initiative Consortium[5]). METS is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library while SCORM is a model that references a set of interrelated technical specifications and guidelines designed to meet the Department of Defense's high-level requirements for Web-based learning content. The TEI-Lite DTD, a low-complexity DTD, was chosen as the metadata schema to identify and describe the digital objects in the Courseware system and their relationships.

TEI is a flexible standard that allows the replication of various free-form structures of Courseware data as closely as possible to the original structure. Courseware was designed for the convenience of professors and students in a face-to-face academic environment and is not optimized for long-term storage or autonomous distance learning. In the future, it may be necessary to require more detailed information from participants in LMS systems, but capturing that information will have a convenience cost that will either result in reduced participation or require tools to be built that will automate the creation process.[6]

[5] TEI is an independent organization supporting an international and interdisciplinary standard to handle a wide variety of texts and text-processing work for online research and teaching, using an encoding scheme to include a high level of detail that would aloud the production of new documents as well as encoding of existing one while usable with a wide range of existing XML software.

[6] The reason GSD chose TEI for this initial assessment and not METS, SCORM, or something similar was because of the content being examined by this project. While there are digital objects in the Courseware data and some of the data might be usable for distance learning, the data cannot stand alone as distinct objects outside the course and only provide meaningful value when evaluated as part of the whole course curriculum. METS or some other standard may still be used to supplement the limited metadata currently gathered, including any original element additions to the base TEI-Lite DTD, as additional metadata will enhance the preservation process for these objects. However, much of the requested metadata is not part of the stored data in the system and may need to be inferred, especially since the development of the Courseware system is on permanent hold. We are aware that TEI is not an ideal solution and a better one may exist for storage of the curricular website information, but in our limited time we did not find it.

A variety of file types were identified in the inventory and some may present future challenges and difficulties.[7] For example, various file formats used for CAD (Computer-aided design) were discovered during the GSD inventory. The programs that read and write these formats are unlikely to be available in the future. This led to a search for an independently supported standard format to store the data. The SVG (Scalable Vector Graphics) format, an XML-based solution for storing visual data in a resolution-independent manner, was the best standard currently available for CAD data. An off-the-shelf application was chosen to convert the CAD data to SVG. The software not only converts the format but also kept some of the associated metadata using XML, XSLT, and SVG with built-in layer controls and navigation tools, plus markup/redlining tools. For GIS (Geographic Information System) files, however, the initial testing indicates that vital interactive layers were either rendered useless or lost in the conversion process to SVG, so this might not be an ideal solution. The storage of the original files in addition to the SVG files is recommended so that a better solution can be adopted if one becomes available while the original files are still readable.

In the GSD case study it was logical to take advantage of all existing administrative information and descriptive references used by the database to supplement core course information. By gathering this comprehensive information now, filters can be created later to selectively extract desired information for future archival projects.

Lessons learned

In the course of this project, technical specialists were able to create a model that could allow for the long-term preservation of the materials outside the Courseware LMS. Perhaps the most important lesson learned during this study phase is the need for established policies, both University-wide and local, *before* any implementation begins. Established policies can serve as a preservation framework for decision-making over time, especially to address questions pertaining to intellectual property and for communicating archival expectations up front. The following conclusions were drawn from the GSD case study:

- For the archiving of LMS materials to be scalable and fully successful, administrative policies and technical specifications designed to support preservation and long-term access must be established at the outset of system design.
- Appropriate metadata is vital to ensure the durability of valuable digital resources. Detailed metadata serves as the connecting link as well as the retrieval and management vehicle for all the components that make up a single course.
- Archivists, metadata analysts, and technical specialists must be involved early in the design stages of system and application development.

[7] 11 different file media types encompassing 51 specific file formats were identified on the active web directories at the Graduate School of Design. The inventory took approximately 80 hours to complete.

Media type	*Format (by file extension)*
CAD	dcd, dwg, dxf, efc, max, mcd
3D	3ds, fmz, wrl
GIS	apr, aux, db, dbf, dir, met, mperr, mxd, shp, shx
Images	ai, bmp, eps, gif, jpg, pct, plt, png, psd, tif
Text	pdf, ppt, ps, qxd, rtf, txt
Mark-up	css, htm, js, xml
Animation	fla, swf
Audio	au, mid, mp3, ram, wav
Video	avi, mov, mpg
Spreadsheet	xls
Compressed	zip

GSD outline for policy decisions

Included here are some key points identified during the GSD case study that will be shared locally and could be applied to other digital materials as well. Addressing and defining some of these points prior to any implementation process will help the GSD fulfill its archival goals and move toward a proactive approach to preserving its history of web-based course materials rather than reactively dealing with specific issues as they present themselves.

Important factors for the GSD to consider:

12) Define the school's archival mission and goals.
13) Identify key players, stakeholders, and local resource providers. Know what service they can and cannot provide.
14) Compile an inventory of information to preserve.
    a. Types of information:
        i. Faculty-created material
        ii. Student-created material
        iii. Co-mingled student and faculty material
        iv. Externally owned material (copyrighted material like journal articles or excerpts from books that may be part of class reserve readings)
        v. Administrative and management data (enrollment information, interactive quizzes, access/usage logs, etc.)
        vi. Other miscellaneous institutionally generated materials (librarian-created guides or documentation on computing facilities prepared by local IT group)
    b. File format types (example from the Design School):
        i. text file types, including HTML
        ii. still image file types
        iii. CAD (computer-aided design) file types
        iv. GIS (Geographic Information Systems) file types
        v. multimedia file types
    c. Administrative metadata:
        i. Content ownership
        ii. Author(s)
        iii. Creation dates
        iv. Content descriptions
    d. Technical metadata
        i. File format types
        ii. Applications used
        iii. Version of the applications
        iv. Size, orientations, and resolutions
        v. Interrelationships
    e. Use and appraisal metadata:
        i. Importance of content
        ii. Retention periods
        iii. Access permissions/audience
    f. Storage and location of metadata (image of CPU, backup tapes, floppies, hard drive)
15) Develop appropriate high-level organizational policies and objectives, ensuring that they are clear and well communicated.
    a. Because there are many types of information and formats, established policies can serve as a preservation framework for decision-making choices over time, particularly for questions pertaining to ownership and use of student/faculty work.
    b. "The primary goals of policy making would be to protect and empower the faculty and the institution by making sure that questions involving privacy and rights are managed appropriately." - Clifford Lynch, Fall 2002 CNI Project Briefing
16) Address intellectual rights/copyright issues.

   a. Student- and faculty-created material
   b. Externally owned material
17) Develop appraisal and retention protocols, and set a time frame for determining when information becomes frozen as a record vs. remaining a living document.
   a. Systems to house information
   b. Systems to permit access during period as living document
   c. Recovery and backup strategies
18) Address access rights to information and under what circumstances access would be granted.
   a. Post-course faculty access
   b. Post-course student access
   c. Re-use of courses
   d. Post-course independent access
19) Identify and categorize information for long-term preservation and information in active record-keeping systems that might only need to be kept for a relatively short time-horizon along with information that has no long-term historical value for the organization.
20) Develop a process to catalog the information thoroughly, giving particular importance to its provenance and transformations of the material.
21) For a given digital resource, address whether its original presentation form (the literal content) or only its intellectual content must be preserved.
   a. For example, it is common at the GSD for information related to a single project to exist partly in traditional paper format, partly in an email box, word-processing file, database tables, and multiple image files generated in a variety of formats, platforms, and programs.
   b. Because of the hybrid nature of information produced, one challenge is to design a suitable solution for the school.
      For example, Geographic Information System (GIS) produced maps, CAD and 3-D images converted and flatted into a two-dimensional image may not be suitable substitutes for the originals. Converted files may lack elements or other critical information vital in the original form.
22) Address organizational responsibilities, ongoing maintenance, and resources.
   Because maintaining archived materials over time will require ongoing investment and attention, addressing economics could very likely further narrow what resources would be deemed as valuable and identified for long-term preservation.

Scope of Courseware system and estimated resources required for case study

- Courses examined (Fall 1999 through Spring 2003): 1443
- Size of generated course data: 8298510 bytes (7M 936K 14B)
- Uploaded files: 4836
- Edited application files: 48
- Courseware modules/tools in original application: 11
- Time commitments (estimated):
    - Learning existing system infrastructure: 4 weeks
    - Learning new technologies: 2 weeks
    - Learning about possible solutions for data storage: 4 weeks
    - Editing the front end of existing application: 15 weeks
    - Writing code to access and manipulate generated data: 9 weeks
    - Documentation: 4 weeks
- Skills set:
    - Significant knowledge of (or willingness to learn) language in which the application is written. In the case at the GSD, this is Cold Fusion.
    - SQL, as part of work with the existing application. In theory, this would not be necessary to alter an application that made no use of SQL.
    - XML and HTML, to convert an existing HTML front-end into XML.
    - PERL, for use in accessing and manipulating data generated by the web site. Similar work could have been done in another general-purpose scripting language, such as Python.
    - XSLT, used to transform XML data.

Acronyms used in the GSD case study

CAD	Computer-aided design
DRS	Digital Repository Service
DTD	Document Type Definition
GIF	Graphics Interchange Format
GIS	Geographic Information System
GSD	Graduate School of Design
HTML	Hypertext Markup Language
IT	Information Technology
JPEG	Joint Photographic Experts Group
LMS	Learning Management System
METS	Metadata Encoding & Transmission Standard
Perl	Practical Extraction and Report Language
SCORM	Sharable Content Object Reference Model
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TEI	Text Encoding Initiative
TIFF	Tag Image File Format
XML	Extensible Markup Language