



Texas Center for Digital Knowledge  
University of North Texas  
<http://www.txcdk.org>

## **The Texas Course Redesign Learning Object Repository: Research and Development for a Production System**

**A Proposal to  
The Texas Higher Education Coordinating Board**

**In Response to  
Texas Course Redesign Project:  
Phase III Full Course Redesign  
Request for Proposal**

**Submitted by**

**Dr. William E. Moen, Principal Investigator**

<wemoen@unt.edu>

Interim Director

Texas Center for Digital Knowledge  
University of North Texas

**Vandana Singh, Co-Principal Investigator**

Texas Center for Digital Knowledge  
University of North Texas

June 29, 2007

**Table of Contents**

1. Introduction ..... 1

2. Challenges and Opportunity Areas ..... 1

    2.1. Technical Development ..... 1

    2.2. Usability Testing ..... 2

    2.3. Operation of the LOR ..... 2

    2.4. Intellectual Property, Licensing, and Related Policies ..... 3

    2.5. Populating the LOR with Course Content from the Texas Course Redesign Project ..... 3

    2.6. Training for and Marketing of the LOR ..... 4

3. Project Goal and Objectives ..... 4

4. Project Design and Plan ..... 5

    Work Area A: Project Management, Documentation, and Communications ..... 5

    Work Area B: Technical Development and Enhancements ..... 5

    Work Area C: Users and Usability ..... 6

    Work Area D: Implementation and Operation ..... 8

    Work Area E: Policies and Procedures ..... 9

    Work Area F: Training and Marketing ..... 10

    Work Area G: Assessment and Evaluation ..... 11

    Summary ..... 11

5. Project Management and Timelines ..... 12

6. Project Budget ..... 12

7. Project Deliverables ..... 14

8. Impact and Benefits of the Project ..... 14

9. Sustainability and Dissemination ..... 14

10. Qualifications and Readiness of the Project Team and Institution ..... 15

11. Proposal Summary ..... 16

References ..... 18

Appendix A: Letter of Support from Texas Digital Library ..... 19

Appendix B: Automating the Generation of Metadata for Learning Objects ..... 22

Appendix C: Brief Curriculum Vitae of Investigators and Researcher ..... 23

Appendix D: Budget ..... 30

---

## 1. Introduction

The Texas Center for Digital Knowledge (TxCDK) at the University of North Texas (UNT) received a grant from the Texas Higher Education Coordinating Board (THECB) under the Texas Course Redesign Project: Phase II Fast Track program to develop a proof-of-concept application of a learning object repository (LOR). Work on that application is underway and will be completed by August 31, 2007. The goal of the awarded project is to demonstrate a proof-of-concept learning object repository's capability to leverage the course content developed through the Texas Course Redesign Project for reuse and repurposing. In this proposal, we refer to the awarded project as Stage 1. The proposed project's goal is to lay a solid technical, organizational, and policy foundation for what we refer to as a near-production ready Texas Course Redesign Learning Object Repository.

This new proposal describes research and development activities to move the proof-of-concept learning object repository forward towards a production level system. The proposed work is multi-faceted with specific attention to key areas including: technical development and enhancements; policy and procedure development related to intellectual property rights and licensing; research into user requirements and usability testing to inform the next version of the LOR; training and marketing; and assessment and evaluation. In addition, substantial content will be added to the LOR from courses being redesigned as part of the Texas Course Redesign Project. Successful inclusion of course content will require THECB to facilitate cooperation communication with those in charge of the course redesign to ensure the content creators understand any requirements for submitting the course content into the LOR and providing TxCDK with information related to the course content needed for metadata creation. The major deliverable from the proposed project will be a near-production ready LOR implemented at the Texas Digital Library, populated with learning objects (LOs) from 4-6 of the courses being redesigned as part of the Texas Course Redesign Project.

## 2. Challenges and Opportunity Areas

A learning object repository (LOR) system must provide the functionality to support both administrative and end-user services. Stage 1 is demonstrating basic repository functionality. An operational, production-level system faces issues that range from technical to policy to training and marketing. Stage 1 work is also highlighting questions that will need to be addressed to move forward to a production-level system providing value-added services for the intended users of the LOR. We briefly describe the categories of challenges and opportunities and/or potential solutions in this section. Subsequent sections of the proposal provide details on how the proposed project will address these challenges and opportunities.

### 2.1. Technical Development

Stage 1 implementation will yield a functional LOR that addresses metadata, submission processes, and end user searching. Near the end of Stage 1 work, we will carry out assessments of the proof-of-concept LOR (as described in the Stage 1 proposal). Those assessments may point to areas for further development.

Stage 1 did not focus on user interface considerations for either submission of learning objects into the LOR or end-user interaction with the LOR. Recent developments by Texas A&M's Digital Initiatives has resulted in new XML-based tool for DSpace (the repository application being used in Stage 1) called DSpace Manakin, a user interface developed by Texas A&M University which supports major improvements for the DSpace system (Texas A&M Libraries, 2007). DSpace Manakin provides new opportunities for customization of the user interface.

The metadata records describing the LOs in Stage 1 result from manual creation processes, which are both time consuming and expensive. For scalability and to improve efficiency in metadata creation, automatic metadata generation offers new opportunities. Natural language processing (NLP) can be used to improve search and retrieval and browsing by extracting keywords from the LOs, identifying key concepts and placing those in relation to other concepts, and automatic summarization of LO content.

## 2.2. Usability Testing

Although Stage 1 assessment of the proof-of-concept LOR included some user testing, it did not include research into user requirements nor systematic usability testing (in part because the Stage 1 focus did not include intensive customization of the LOR user interfaces). As noted above DSpace Manakin provides a tool for easier and more flexible user interface customization. Users require web-based applications to be easy to use and navigate. Success of the LOR may ultimately depend on the way the functionality and services of the LOR are presented and available for the users.

To ensure quality of the interface customized to the target audience usability engineering techniques can be used to inform the design of the interfaces. At the heart of this effort is the recognition of usability as a quality attribute (Nielsen, 1993). Usability is defined by five quality components:

- **Learnability:** How easy is it for users to accomplish basic tasks the first time they encounter the design?
- **Efficiency:** Once users have learned the design, how quickly can they perform tasks?
- **Memorability:** When users return to the design after a period of not using it, how easily can they reestablish proficiency?
- **Errors:** How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- **Satisfaction:** How pleasant is it to use the design?

Usability engineering comprises usability testing characterized by Rubin (1994) as “the process that employs participants who are representative of the target population to evaluate the degree to which a product meets specific usability criteria.” Other usability engineering techniques which do not require users as part of process will also be used (mainly heuristic evaluation, and cognitive walkthroughs). Based on the input from these testing techniques, an iterative development process will be used for customizing the LOR interface.

## 2.3. Operation of the LOR

According to THECB staff, THECB does not plan to operate and maintain the Texas Course Redesign LOR. Therefore, it is critical to plan for an operational home where the LOR will be hosted and will be administratively and operationally maintained. In discussions with THECB staff, we proposed that the Texas Digital Library provides a potential home for the LOR.

The Texas Digital Library “provides a digital infrastructure for the scholarly activities of Texas universities. On university campuses throughout Texas, an enormous amount of intellectual capital exists that is not readily available to faculty, staff, and students throughout the State. TDL will serve as the center of excellence for the creation, curation, and preservation of digital scholarly information for the State” (Texas Digital Library, 2007). The Texas Digital Library (TDL) was founded and is collectively managed by four Association of Research Libraries (ARL) universities: The University of Texas, Texas A&M University, The University of Houston, and Texas Tech University. Currently 14 public institutions of higher education participate in the TDL. The University of North Texas is a Tier 2 Associate Member. (Note: The principal investigator for the proposed project is chair of the TDL’s Metadata Working Group.)

Included in planned collections of digital scholarly information for the TDL are learning objects, and TDL plans to implement a LOR in the future. TDL has committed to using DSpace for its repository applications. The Co-Directors of the TDL have expressed strong interest in working with the Principal Investigator to host the next version of the Texas Course Redesign LOR. Appendix A contains a letter of support and commitment from the TDL Co-Directors. Work will be needed to migrate the application from UNT’s TxCDK to TDL. In addition, agreement on responsibilities for ongoing administration and operation will be necessary.

## **2.4. Intellectual Property, Licensing, and Related Policies**

The Texas Course Redesign Project intends to make available to all public institutions of higher education the course content in the redesigned courses. In effect, THECB will own the intellectual property of these redesigned courses in whole or in partnership with the recipients of THECB funds for redesigning these courses. However, activities in Stage 1 are highlighting a more complex situation that will require attention to intellectual property (IP) and licensing questions.

In Stage 1, THECB requested the proof-of-concept LOR include content from an English Composition course being redesigned by the Dallas Community College (LeCroy Center) and Texas Tech. This course includes video components that were created and produced by the LeCroy Center without THECB resources. The LeCroy Center is using THECB funding to redesign this course and develop new interactivity components for the course, but the redesigned course includes content that was not funded by THECB. This may not be the only instance where redesigned courses include legacy content in addition to THECB-funded content. There may be complicated IP rights and licensing involved for use of the THECB redesigned courses by Texas institutions of higher education.

Stage 1 is demonstrating how course content can be decomposed into discrete reusable and repurposable learning objects. The potential for reuse and repurposing raises interesting questions regarding the rights institutions will be given to modify THECB-funded LOs and the additional intellectual property aspects to modified THECB LOs. Licensing agreements are needed that articulate what types of uses are allowed and the disposition of modified LOs once they are downloaded from the LOR.

The content of the Texas Course Redesign LOR may have value to institutions other than the intended Texas public higher education institutions, such as private institutions in Texas, secondary education schools, as well educational and other organizations outside of Texas. THECB may consider opportunities for licensing learning objects to such institutions and organizations.

Policy decisions and legal/licensing solutions are needed, and once completed, the LOR's technical capabilities need to be enhanced or configured to implement such policies and licenses through access controls, permissions, or other mechanisms.

Another area of policy decisions relates to the operation of the LOR. For example, one area of operational policy will need to address who will have the right and obligation to submit content into the LOR. In the digital repository environment there are different models for submission of content – from very centralized and controlled models to distributed and relatively open submission processes. Another aspect of submission is the creation of metadata records to describe the LOs. These records play a critical role in finding, identifying, selecting, and accessing/obtaining the LOs from the repository. General operational policies for the LOR are necessary.

## **2.5. Populating the LOR with Course Content from the Texas Course Redesign Project**

The value and utility of the Texas Course Redesign LOR increase as the quantity and diversity of learning objects in the LOR increase. The Stage1 proof-of-concept LOR includes sufficient content (primarily from two courses in different academic disciplines) to demonstrate aspects of the LOR as described in the earlier proposal. Ingesting new content into the next version of the LOR will further prove out the application as well as the procedures for submitting content and metadata creation being developed in Stage 1.

Working with new content from more disciplines raise important questions of disciplinary terminology to use in metadata records. In addition, different disciplines and different types of courses may have little in common with the structural organization and component parts of the courses being used in Stage 1. New questions related to the appropriate granularity for learning objects in different types of courses are likely to emerge.

## 2.6. Training for and Marketing of the LOR

THECB will not want the Texas Course Redesign LOR to be the best kept secret in Texas. Having a high quality product, which is what the LOR can be, is the foundation on which an active marketing effort will be based. The marketing effort needs to reach the intended “customer base” and the development of a plan to guide the marketing effort is the necessary first step.

Stage 1 work and the work described in this document are intended to yield an easy-to-use discovery and access mechanism for high quality learning objects. Good design and functionality, tested and improved through the usability assessments listed previously, may dramatically reduce the need for training of end-users who are searching for learning objects. Similarly, good interface design should make it easy for content creators to submit learning objects into the LOR (but it should be noted that THECB will need to develop policies related to submission of content, metadata creation, etc. – see 2.4 above). Appropriate training and/or training materials should be considered, and the nature and extent of those materials can be informed by the outcomes of usability testing. Online training modules and just-in-time context sensitive help on using the LOR may be preferred methods for instructing potential users.

A good product is easier to sell than a deficient one, but a good product can only go so far in selling itself. Marketing is made easier with a web-based application that has the hallmarks of good design and has addressed usability criteria. Ensuring that appropriate training and training materials are available will assist the user becoming comfortable and efficient in using the system. And of course, sufficient high quality course content in the form of reusable and repurposable learning objects is the core value the Texas Course Redesign LOR will offer to public institutions of higher education in Texas.

## 3. Project Goal and Objectives

The proposed project’s goal is to lay a solid technical, organizational, and policy foundation for a near-production ready Texas Course Redesign Learning Object Repository. This requires both research and development activities as reflected in the following project objectives:

1. Develop and implement a near-production ready LOR application reflecting enhancements based on Stage 1 outcome and insights.
2. Design and customize the LOR user interfaces (end user interfaces and content submitter interfaces) and assess those interfaces using accepted practices and criteria for usability.
3. Examine and implement where appropriate computational methods and processes to assist in automatic metadata generation.
4. Develop, in consultation with THECB staff, draft operational and administrative procedures for a production-level LOR system.
5. Explore opportunities for the Texas Digital Library (TDL) to serve as host for the LOR by implementing the near-production ready LOR on systems operated by the TDL.
6. Initiate, coordinate, and lead discussions, in consultation with THECB staff, with potential stakeholders of the Texas Course Redesign Learning Object Repository, related to intellectual property rights, licensing, and related policies, resulting in a draft of such policies for consideration of THECB and the stakeholders.
7. Ingest learning objects into near-production ready LOR drawing on course content from 4-6 of the redesigned courses (as available for the proposed project).
8. Develop an appropriate training program and/or training materials to assist end users and content submitters in using the LOR.
9. Develop a draft plan, in consultation with THECB, to market the LOR to intended users and other potential customers.
10. Assess the achievement of project goal and objectives, and develop a plan for ongoing evaluation of the LOR when it is in production.

These are wide-ranging objectives, but all focused on laying the groundwork for a production-level LOR system and set of services. Several of these objectives entail research activities into understanding user requirements, user behaviors, and user assessments of the usability of the LOR, as well as research and development activities to automate some aspects of metadata creation.

## 4. Project Design and Plan

The proposed project consists of major work areas that encompass the project objectives. This section describes each of the work areas, and in some cases, specific components of the work area are detailed.

### Work Area A: Project Management, Documentation, and Communications

This work area addresses all activities to manage and guide the project to a successful conclusion, and includes documentation of project activities and responding to reporting requirements.

The success of any project depends on active management of the human, technical, and financial resources available for the project. Managing for success includes the creation of a detailed project plan to guide all project work. Upon award of the grant, the Principal Investigator will develop a detailed project plan.

The Principal Investigator will also establish a project website for dissemination of information about the LOR. The proposed website focuses on project information and is not intended to substitute for the website currently under development by THECB for its course redesign project. When that site is available, we will request a link to our project website. We believe our proposed work will benefit from broader community awareness and the website can serve in part to nurture such awareness.

#### Key Activities

- Develop a detailed project plan that identifies all work areas' activities, deliverables, timelines, and staffing
- Prepare appropriate technical documentation for all repository components
- Create interim and final reports as required by THECB.
- Establish project website for public dissemination of project information
- Active management of human, technical and financial resources.

### Work Area B: Technical Development and Enhancements

Using Stage 1 deliverables as a starting point, this work area focuses on further technical development of the LOR along with research and development to assist in metadata creation and improved search and retrieval. In addition, a major focus will be on customization and refinement of user interfaces for public end users and also content creators who may be submitting materials into the LOR. Several key components of this work area are described below. The result of Work Area B will be a new version of the LOR that reflects technical improvements and enhancements.

**B.1. Stage 1 Assessments Integration:** Stage 1 work involves two levels of system assessment: functional and user. The results of these assessments will be reviewed and incorporated in a new version of the LOR that will also include enhancements from the following two activities.

**B.2. Manakin implementation and user interface customization:** The Digital Initiatives group within the Texas A&M University Libraries has developed DSpace Manakin, a XML-based user interface tool for use with DSpace (the open source repository application being used in Stage 1, which will continue to be used in the proposed project). Manakin provides new opportunities for customizing DSpace and adding functionality to repository applications. In the context of the LOR, Manakin will enable unique look-and-feel for disciplinary areas and specific courses. TxCDK began exploration of DSpace Manakin in Spring 2007. Based on user research results from Work Area C. Users and Usability, we will use Manakin to customize user interfaces which will be subject to Work Area C usability testing. We have discussed the use of Manakin for the LOR with John Leggett, Director of the Digital Initiatives group and Associate Dean, Texas A&M University Libraries, and he has agreed to provide consultation and assistance in training project staff on Manakin and its implementation in the LOR.

#### Key Activities

- Install DSpace Manakin in a test environment

- Consult with and receive training from Texas A&M technical staff on the functions and operations of the DSpace Manakin user interface
- Customize the LOR interfaces based on results of Work Area C activities
- Investigate additional functionality and services DSpace Manakin can enable in the LOR.

**B.3. Computer-assisted metadata generation for search, retrieval, and browsing:** Metadata ("data about data") is information about objects and their content that can facilitate the use, understanding, and management of collections of data. Although metadata has been traditionally created manually, the metadata community has recognized the challenges with relying on labor-intensive human-generated metadata, given the increasing amount of digital resources being produced and needing to be described. The problem has been referred to as the metadata bottleneck.

Natural language processing (NLP) is being used to assist in breaking this bottleneck by bringing the power of computer processing to the generation of metadata. We will examine how machine-assisted metadata creation can complement and support human metadata creation. Dr. Rada Mihalcea, Assistant Professor in UNT's Department of Computer Science and Engineering and a member of the project team, will coordinate the work on metadata generation for the LOR content. Appendix B contains examples of output from several potentially useful NLP algorithms. Dr. Mihalcea has developed keyword extraction techniques that will be refined to address learning objects in the LOR to improve query/retrieval results. In addition, Dr. Mihalcea will investigate the opportunities to identify related concepts (beyond the extracted keywords) addressed by LOs even when they are not explicitly mentioned in the LO (e.g., identifying the concept of "history" in a course lesson that covers the topic of World War II).

The availability of automatically extracted keywords and associated conceptual labels may improve the discovery of potentially relevant LOs by instructors in disciplines outside of the discipline for which the course the LO was created for. Additionally, the keywords extracted from LOs and the related concepts can be organized in ontological structures that can be used as an effective navigation tool through LOs. Ontology-like structures provide support for browsing through specification (e.g., suggest LOs related to the more specific concepts of "Battle of the Atlantic" and "Liberation of Western Europe" in response to a query for "World War II") or generalization (e.g., suggest LOs relevant to the more general concept "World Wars" or "Historical Events" in response to a search for "World War II").

Another area of research that will be explored is automatic summarization of textual LOs, which could result in an efficient method for describing a LO, and the description can be included in the metadata record. Summaries can be used to improve the search and retrieval of LOs, or as a means to facilitate the browsing experience of LOR users by providing brief summaries of the LOs of interest.

#### **Key Activities**

- Develop keyword extraction methods appropriate for LOs and implement a tool for keyword extraction and inserting those automatically extracted terms into the metadata records.
- Test methods and tools for identification of conceptual labels associated to LOs
- Conduct research for building ontologies based on the identified keywords and conceptual labels for better search and retrieval of LOs and for an improved navigation experience through the LOR
- Conduct a pilot-study to identify the potential for automatic summarization of LO content for consideration of use in future LOR development.

#### **Work Area C: Users and Usability**

The Stage 1 proof-of-concept project's proposal stated:

The timeframe for this project does not allow research into user requirements, user behaviors, and user testing of the proof-of-concept system developed during the project. If, based on the results of this project, THECB determines to move forward towards an operational learning object repository, systematic user research is critically important to ensure that the operational system responds to the needs and behaviors of potential users.

To ensure the LOR, when in production meets the needs of its intended audiences, it is critical to have sufficient information about the users, their requirements, vocabularies, and other related information. In addition, the LOR must meet accepted usability criteria to make it easy for users to engage with the application to find, identify, select, and access or acquire relevant learning objects. Work Area C includes key components that involve research about the intended users and the usability of the application. Results from activities in Work Area C will directly inform Work Area B activities.

Vandana Singh, Co-Principal Investigator, is defending her dissertation this summer and joining the UNT faculty in August 2007; she will have primary responsibility for Work Area C activities.

**C.1. User requirements identification:** Many of the intended users of the LOR may have little understanding of what a learning object is, and specifically what the Texas Course Redesign LOR is and what it makes available. There is a need to understand at least the following from the potential users: 1) how they will search and browse the repository for items of interest; 2) the levels of granularity most appropriate for the LOs in terms of these users needs; 3) the specific disciplinary vocabularies that are appropriate in metadata descriptions of discipline-specific LOs; and the critical information needed in a metadata record to support these users to search, identify, select, and access or obtain relevant LOs.

The specific methodologies for collecting this information will be developed as part of the project work in this area; however, the general plan for data collection includes the use of focus groups, user and task analysis and low and medium fidelity prototyping with representatives of the intended users. The focus group will inform user preferences and will aid in developing usability metrics for usability testing. Participants in the focus group sessions will be chosen based on the disciplines for which redesign courses are being created. For example, a group of instructors for U.S. History courses will participate together in the focus group. As a follow up from the focus group, field studies will be conducted for user and task analysis – in this stage the users will be given a task to perform and they will be observed in their context of work. The data collected from this will inform the iterative design of the LOR interface. We will use the Stage 1 proof-of-concept implementation and prototypes to demonstrate what the LOR is and how it works, and then elicit the users suggestions on functions, granularity levels, metadata records, and terminology. Results of this user research will inform technical development and enhancement work in Work Area B.

#### **Key Activities**

- Develop a research design and methods to collect information reflecting the requirements of intended users of the LOR
- Conduct the user research, collect and analyze the data, and compile results
- Identify specific technical and functional requirements based on the results to inform Work Area B activities.

**C.2. Usability assessment:** A key success factor for the LOR is the extent to which the intended users can find, identify, select, and access the LOs stored in the repository. To address this success factor, the LOR application must meet or exceed acceptable usability criteria. We will develop the detailed methodologies and procedures for conducting the usability assessments. Broadly we will be doing user-based usability testing by using low and medium fidelity prototypes and field studies and expert based usability testing by heuristic evaluation and cognitive walkthroughs. Areas of the LOR subject to the usability testing include: search and navigation, organization, look-and-feel, metadata records, and results display. We envision an iterative development of the LOR interface based on the user requirement identification and usability testing.

#### **Key Activities**

- Develop usability testing methods and procedures specific to the LOR application
- Identify testers who are potential members of the LOR's intended user groups
- Conduct the usability testing, collect and analyze the data, and compile results
- Identify specific changes to the application based on the results to inform Work Area B activities.

## Work Area D: Implementation and Operation

Work Areas B and C will result in a new version of the LOR that reflects user requirements, usability assessments, and other enhancements. Work Area D addresses the implementation of the LOR at the Texas Digital Library (TDL) and the implementation of operational procedures resulting from Work Area E. Policies and Procedures for administering the LOR. A major component of this work area will involve submitting content into the LOR from various courses developed for the Texas Course Redesign Project. The key components of this work area are described in the following sections.

**D.1. Implement the LOR in the Texas Digital Library:** As indicated in Section 2.3, the directors of the TDL have indicated their interest and support for hosting the next version of the LOR in a test environment in the TDL (see Appendix A). We will work with the technical staff at TDL to prepare an instance of DSpace (the repository application software used by TxCDK and by TDL) and then migrate/implement the new version of the LOR on TDL-hosted computing resources (these resources are currently located at the University of Texas at Austin campus). Once implemented, project staff will be responsible for testing the TDL-hosted instance of the LOR. TDL will provide core systems administration functions on the hosted LOR.

### Key Activities

- Engage with TDL technical staff to set up a DSpace instance hosted by TDL
- Migrate/implement the new version of the LOR from the TxCDK server to the TDL server, which will provide the LOR in a controlled access, test environment
- Fully test the TDL-hosted implementation to prepare it for submission of course content.

**D.2. Implement the operational policies and procedures:** Work Area E deliverables include operational policies and procedures. These policies and procedures will govern the TDL-hosted LOR. These policies may address controlled access to the repository for submission of course content, metadata creation, editing, and other tasks. To the extent these policies and procedures are enforceable by the software application, the TDL-hosted LOR will be configured appropriately to ensure the policies and procedures are in force.

### Key Activities

- Prior to migrating to TDL, operational policies and procedures enforceable by the software application will be tested and refined in the new version of the LOR
- After migration to TDL, the LOR's configuration will be tested to ensure the policies and procedures are in force.

**D.3. Populate the LOR with content:** THECB anticipates a number of redesigned courses will be completed in a timeframe complementary to this proposed project. Specifically, THECB staff identified the following courses to anticipate for inclusion in the next version of the LOR: 1) a calculus course; 2) an algebra course; 3) two Spanish courses; 4) a developmental English composition course. As with the U.S. History 1 course being used in the Stage 1 proof-of-concept, the project team will work with the creators of these additional courses to identify appropriate levels of granularity at which to store the LOs. User research conducted in Work Area C will yield information about discipline-specific vocabularies needed to represent and describe the LOs. Success in populating the LOR with content from these additional courses will require THECB help in ensuring the content creators work cooperatively with the project team to acquire the course content and to provide needed information about the content. In addition, LOs from the U.S. History 1 course submitted to the Stage 1 proof-of-concept implementation will be migrated to the TDL-hosted LOR.

### Key Activities

- In consultation with THECB staff, the project team will determine the redesigned courses' content to bring into the next version of the LOR
- In cooperation with THECB staff, the project team will communicate with the creators of the redesigned courses to acquire the course content in formats and structure appropriate for submitting into the LOR

- In cooperation with THECB staff, the content creators will provide sufficient contextual and other information about the course content to serve as a basis for the metadata records describing the LOs
- Migrate Stage 1 LOs to the TDL-hosted LOR
- Populate the TDL-hosted LOR with new content from selected redesigned courses.

### **Work Area E: Policies and Procedures**

THECB will need to consider a variety of policies and procedures for the LOR. In particular, LOR policies related to intellectual property (IP) and licensing are critical. Also of interest are policies governing reuse of the LOs, especially regarding modification of LOs originating from the LOR. Finally, the operation of the LOR, wherever is housed, should be governed by policies to ensure transparency and accountability; this includes operating policies and procedures relating to who has rights to submit to the LOR, create metadata, and other facets of a production system. The primary components of this work area are described in the following sections.

**E.1. Intellectual property and licensing policies:** The RFP indicates that THECB holds the intellectual property (IP) rights on course content developed under the program:

The intellectual property developed as part of the TCRP pursuant to this contract is the property of the THECB. The THECB grants to the Award Applicant(s) a license to sell the intellectual property described in this RFP to private institutions of higher education, as well as to other institutions of higher education outside Texas. The profits of such sales will be divided equally between THECB and the Award Applicant(s). The intellectual property described in this RFP will be made available without charge to public institutions of higher education in Texas.

From experience in Stage 1, we are aware that some of the redesigned courses are including content developed previously and without funding from THECB. THECB will need to decide whether content in addition to that developed with THECB funding will be made available via the LOR.

To address these and related issues such as access and use of LOR content by non-public institutions of higher education in Texas and institutions of higher education and others outside of Texas, we propose to facilitate and coordinate discussions among the key stakeholders (e.g., THECB, content creators and their parent institutions, commercial companies) to develop appropriate and consensus-based policies related to IP and licensing. To carry this out, we will identify an appropriate expert in the areas of IP and licensing to serve as a consultant on this work and to assist in developing draft policies.

Dr. Philip Turner, UNT's Vice Provost for Learning Enhancement, brings valuable experience in the areas of IP and licensing policies for course content and has agreed to contribute to this work. He will assist in identifying a qualified consultant to work with the project team (described below), and will serve as an expert advisor on the development of appropriate policies.

#### **Key Activities**

- Develop an appropriate methodology for building consensus on IP and licensing policies for the LOR
- Identify consultant to serve as facilitator for discussions among stakeholders
- Develop draft IP and licensing policies for consideration by stakeholders
- Implement the resulting IP and licensing policies, as appropriate, in the LOR.

**E.2. Reuse, repurposing, and modification of LOs:** LOs may be reused in their original form, or they may be subject to modification and customization for subsequent use. This presents another policy arena of IP and licensing issues. For example, if an individual or institution modifies and enhances a LO from the LOR, will that person or organization be required to resubmit the revised LO back into the LOR for others to use? Will there be in place a process to request co-authorship of a modified LO? If so, who has the right /obligation to review and grant co-authorship status? What new IP and/or licensing rights will attach to the modified LO? These and other questions will need to be addressed. Activities as part of this

component will use the IP and licensing consultant referenced above to assist in developing a set of options for consideration of the stakeholders.

**Key Activities**

- Identify options for IP and licensing rights related to modified LOs originating from the LOR
- Develop draft IP and licensing policies for consideration by stakeholders
- Implement the resulting IP and licensing policies for modified LOs, as appropriate, in the LOR.

**E.3. Operational policies and procedures for the LOR:** The IP and licensing policies that will be developed need to be complemented by written operational policies and procedures for the LOR. The operational policies can address a range of questions, such as: who will be able to submit LOs into the repository; and can LOs be removed from the LOR, and if so, under what circumstances? The creation of metadata records describing the LOs is another aspect of the LOR that needs discussion and agreement. The quality of the metadata is key for users searching for and finding relevant materials. One approach is to have digital librarians primarily responsible for metadata creation rather than the content creators and submitters. To address this component of the project, we will develop recommended procedures and policies for consideration by THECB.

**Key Activities**

- Identify key areas for the operation and procedures of the LOR that should be governed by explicit and published policies
- Draft appropriate policies and procedures for consideration of THECB
- Implement the resulting policies and procedures for the LOR.

**Work Area F: Training and Marketing**

Deliverables from the work areas discussed previously are intended to yield a near-production-ready LOR. This work area addresses two complementary components that will assist users in engaging with the LOR and informing the intended audience of the LOR's for their courses. As indicated previously, we are aware that THECB is creating a web presence of its Texas Course Redesign Project. Training and marketing materials developed for the LOR will be housed on that website when it is available.

**F.1. Training:** The design goal for the LOR is to make the application easy to use, and therefore reduce training requirements for the intended users. Responding to user requirements and conducting rigorous usability testing to inform design, navigation, and other facets of the application will support the design goal. However, end users may need assistance in getting started using the system. We propose just-in-time training assistance in the form of an online training module and appropriate contextual and other help pages users can reference as they engage with the system.

The decisions related to policies about submission of materials into the LOR (discussed in E.3. Operational policies and procedures for the LOR) will inform the extent of training materials needed for potential submitters to the LOR. Stage 1 deliverables include a document that describes the submission process and metadata creation. This document will be revised to reflect changes in the next version of the LOR. If appropriate, an online training module will be created for potential submitters to explain the submission process and metadata creation.

**Key Activities:**

- Identify the need for training and information pages for intended end users of the LOR
- Develop an online training module that explains and describes the ways in which intended end users can search, browse, and access the LOR
- Develop appropriate help pages or other contextual help resources for just-in-time use by intended end users as they engage with the system.
- Assess the online training module and help pages with representatives of intended LOR users.

**F.2. Marketing:** The potential benefits of the LOR to Texas institutions of higher education will be realized by ensuring that the intended users are aware of the LOR. This requires active information dissemination

and marketing of the LOR. In collaboration with THECB staff, we will develop an appropriate marketing plan that will use a variety of communication channels to reach the intended users. These channels can include traditional venues such as presentations at appropriate state conferences, press releases, and email. Innovative marketing can also include the use of YouTube (for an example of a recent video marketing UNT Libraries' Portal to Texas History, see: <<http://www.youtube.com/watch?v=rIDx9n4wFb0>>).

**Key Activities:**

- Identify potential methods and communication channels for marketing the LOR
- In collaboration with THECB, determine the scope for a marketing plan
- Develop a draft marketing plan for review by THECB
- Finalize the marketing plan.

**Work Area G: Assessment and Evaluation**

Other work areas have identified either formal or informal methods that will be used for assessments of the next version of the LOR (e.g., usability testing). This work area addresses the evaluation of the success of the project in achieving its goals and objectives, and also preparing a plan for assessment and evaluation of the LOR once it becomes operational.

**G.1. Project evaluation:** The proposed project identifies a set of objectives (see Section 3) and a set of deliverables (see Section 7). The objectives and deliverables serve as a baseline by which the project will be evaluated. The major deliverable will be the near-production ready LOR hosted at the Texas Digital Library, populated with content from approximately 6 redesigned courses. The project evaluation will assess the extent to which the project objectives were met and the promised deliverables were produced in an acceptable form to the THECB.

**G.2. Assessment of the operational LOR:** During the proposed project, a number of assessments will be carried out to inform the design and usability of the LOR. For example usability testing will be carried out to ensure that users of the LOR can easily and effectively find and use relevant LOs. It is important, however, to plan for ongoing assessment of the LOR once it becomes fully operational. We propose to develop an outcomes-based evaluation plan for assessing the operational LOR. In collaboration with THECB, we will identify outcomes, indicators, and measures to be used in assessing the LOR. We will develop an evaluation plan that indicates what, how, and when data will be collected and the appropriate data analysis procedures. This plan will be pilot tested with the near-production ready system to ensure is utility.

**Key Activities:**

- In collaboration with THECB, identify outcomes, indicators, and measures for assessing the LOR
- Develop an outcomes-based evaluation plan for review by THECB, and revise with feedback from THECB
- Conduct a pilot test of the evaluation plan with the near-production ready LOR, and make revisions to the plan, if necessary.

**Summary**

This proposal outlines a multi-faceted plan of work to move the Texas Course Redesign Learning Object Repository forward to a near-production ready system. Accomplishment of the project goal and objectives will be contingent on a number of key success factors including:

- Active project management guided by a detailed project plan and supported by adequate resources
- Collaboration and communication with THECB staff on critical project components such as the development of policies and procedures, the marketing plan, the outcomes-based evaluation plan

- THECB engagement and communication with the organizations and individuals responsible for the redesigned courses to ensure cooperation with the proposed project activities
- Cooperation of the organizations and institutions redesigning courses whose content is to be stored in the LOR, particularly in providing information to the project team about the course content, appropriate levels of granularity of LOs, disciplinary vocabularies, course structure, and related information to assist in describing and storing the LOs.

TxCDK's work on the Stage 1 proof-of-concept learning object repository provides a solid foundation for moving forward with the proposed work. The understanding and experience gained through Stage 1 will inform the next step in developing the Texas Course Redesign Learning Object Repository.

## 5. Project Management and Timelines

The project goal and objectives provide the overall framework for this project. The period of work for the project, per the RFP is September 1, 2007 through August 31, 2008 (12 months). Effective management is critical to achieve the project's stated goal and objectives and to produce the deliverables. Effective management will involve appropriate planning, oversight, and ongoing monitoring to ensure milestones are met and deliverables are produced. Dr. Moen will serve as overall project manager and allocate staff resources to the work areas to ensure continuing and concurrent progress on the multiple work areas. Dr. Moen, Ms. Singh, and Dr. Mihalcea will be responsible for leading efforts in the work areas as indicated in Table 1, which also shows anticipated duration of each work area. More detailed timelines for the components of each work area will be articulated in the detailed project plan.

Dr. Moen is the Principal Investigator on the current THECB-funded project to develop a proof-of-concept learning object repository. He has managed a number of large research and development projects (see Appendix C for brief curriculum vitae), including two National Leadership Grants funded by the U.S. Federal Institute of Museum and Library Services, and a two-year project to design and develop a metasearch implementation for the Library of Texas. As principal investigator, Dr. Moen was responsible for overall project design and management, management of project funds, and staffing. UNT will provide accounting and billing services for the proposed project, and the project will follow all appropriate UNT administrative procedures related to staffing, payment of salaries, travel, and other aspects of the project where expenses are incurred.

## 6. Project Budget

To accomplish the goal and objectives for the proposed project, we are requesting \$XXXX. Appendix D details the expense categories and amount requested.

Dr. Moen, as Principal Investigator, will devote 20% level of effort during the 9-month academic year and 50% level of effort in Summer 2008 to the project.

Ms. Singh, as Co-Principal Investigator, will devote 20% level of effort during the 9-month academic year and 25% level of effort in Summer 2008 to the project.

Support for Dr. Rada Mihalcea and Dr. Ron Carriveau, natural language processing expert and learning assessments experts respectively, is included in the budget. Dr. Mihalcea will be supported for 1.5 months level of effort (and will work with a research assistant hired for 12 months). Dr. Carriveau will be supported for a 10% level of effort for six months.

The majority of the requested funds will provide wages and benefits for UNT graduate students who will serve as research assistants and programmers on the project. These students bring expertise and domain knowledge in areas of information technology, information science, repositories, metadata, history, and course design. The project provides a vital opportunity for these students to further develop their knowledge and skills, and will assist in their development as the next generation of higher education personnel for whom online courseware, technology, and repositories will be their common tools. The number of student research assistants is justified by the range of work outlined in this proposal.

Table 1. Project Timeline

Work Area	Lead	Sep 2007	Oct 2007	Nov 2007	Dec 2007	Jan 2008	Feb 2008	Mar 2008	Apr 2008	May 2008	Jun 2008	Jul 2008	Aug 2008
Work Area A: Project Management, Documentation, and Communication	Moen												
Work Area B: Technical Development and Enhancements	Moen & Mihalcea												
Work Area C: Users and Usability	Singh												
Work Area D: Implementation and Operation	Moen												
Work Area E: Policies and Procedures	Moen												
Work Area F: Training and Marketing	Singh												
Work Area G: Project Assessment and Evaluation	Moen & Singh												

The budget also contains a small amount for travel expenses for the Principal Investigator and Co-Principal Investigator to attend meetings with THECB staff related to the project, and for the Principal Investigator and programmer to travel to Texas A&M for training and consultation on the use of DSpace Manakin. Consulting services for DSpace Manakin, migrating/implementing the LOR at TDL, and for the development of intellectual property and licensing policies are included in the proposed budget.

UNT will provide various resources at no cost to the project, including but not limited to: workspace for project team members; computing resources; network access; telecommunications; office supplies; copying; and delivery services.

## 7. Project Deliverables

The deliverables resulting from the proposed project relate directly to the project objectives. The deliverables take the forms of technology applications, technical and policy documents, as well as required reports to THECB, specifically:

- The implementation at the Texas Digital Library of a tested and near-production ready LOR application populated with learning objects from 4-6 redesigned courses
- Natural language processing tools for automatic metadata generation and insertion of resulting metadata into LOR metadata records
- Written policies and procedures for the operation and administration of the LOR
- Written policies related to intellectual property, licensing and related topics that reflect the interests of THECB and other key stakeholders
- Appropriate training materials for users of the LOR
- A marketing plan for the LOR
- An evaluation of the success of the project in achieving its goal and objective
- An outcomes-based evaluation plan for ongoing assessment and evaluation of the LOR once it is in full production
- Interim and final reports required by THECB.

## 8. Impact and Benefits of the Project

The Texas Course Redesign Learning Object Repository, when operational, has the potential to critically impact Texas institutions of higher education, instructors, and students. The proposed project will move this LOR to a near-production ready stage. The LOR will support broad access to the redesigned course content and support not only reusing but also repurposing the learning objects. The user research and usability testing carried out in the proposed work will help to develop awareness and expectations of the intended users of the LOR. Those users will benefit by having the LOR reflect their requirements for functionality and usability.

Implementing the next version LOR at the Texas Digital Library will also heighten the impact and raise the visibility of the LOR since TDL is intended to be a central location of finding and accessing scholarly resources produced by Texas institutions of higher education. THECB will be leveraging the resources available at TDL for hosting the LOR.

As with Stage 1 work, individual institutions that are considering the development of in-house learning object repositories will benefit from the outcome of the proposed project by having access to the technical and other documentation produced by the project. More importantly, the policies related to intellectual property and licensing can serve as models for individual institutions and the locally-created learning objects. In addition, the near-production ready system will provide a model for others to consider in terms of usable design, metadata generation, and other enhancements developed through this project.

## 9. Sustainability and Dissemination

The near-production ready LOR that will be delivered by this project will be accompanied by a set of operational and administrative policies and procedures. Those will be developed in collaboration with

THECB and will identify resources needed for the continuing operation and maintenance of the LOR. The partnership being established with the Texas Digital Library for hosting the near-production ready LOR lays the groundwork for cost-effective operational home for the LOR. As part of the project, TxCDK will work with TDL to identify resources needed for the operational LOR once this project is completed.

The project's development of an outcomes-based evaluation plan can be used to collect data that will inform future enhancements and services that will ensure the LOR is meeting the needs of THECB and the intended users. Maintaining and enhancing the LOR can assist in ensuring long-term viability and sustainability.

The project will document all aspects the technical and policy work carried out and these documents will be publicly available (with permission of THECB) on a project website. This will provide ongoing dissemination of the results of the project and serve as a point of information about learning object repositories. The marketing plan deliverable will enable the THECB to disseminate information and build awareness about the LOR.

Because of the critical research activities related to users, usability testing, metadata generation, policy development and other facets of the proposed work, we look forward to preparing more articles and conference presentations that will also disseminate project findings. In addition, the Principal Investigator will be available to THECB to present information about the project a meetings and conferences that THECB suggests.

## 10. Qualifications and Readiness of the Project Team and Institution

A team of specialists in learning object repositories, automatic metadata generation, usability testing, evaluation, and other relevant areas will work collaboratively to achieve the project goal and objectives, and to produce the identified deliverables. The project team includes the investigators and appropriate researchers and research assistants. Dr. Moen and Ms. Singh will serve as Principal and Co-Principal Investigator (Appendix C contains abbreviated curriculum vitae for the investigators and the natural language processing expert).

**Dr. William E. Moen**, Interim Director, Texas Center for Digital Knowledge (TxCDK), will serve as Principal Investigator. Dr. Moen has designed and implemented several repository projects and is an expert in metadata schemes and their application. He also has expertise in outcomes-based and other evaluation approaches. He will serve as overall project manager and will assemble a group of Master's and Ph.D. students from the School of Library and Information Sciences and the Department of Computer Science and Engineering to carry out work described in this proposal.

**Ms. Vandana Singh**, is completing her Ph.D. at the University of Illinois and is joining UNT's School of Library and Information Sciences in Fall 2007. In addition to her Ph.D., she has Masters degrees in Computer Science and Knowledge Management Systems. Her academic preparation and involvement in various research projects involving users bring the needed expertise for the components of the project she will lead, namely, identifying user requirements and usability testing components. In addition, she will lead the development of training materials.

**Dr. Rada Mihalcea**, Assistant Professor, Department of Computer Science and Engineering, UNT, is an expert in natural language processing, and will lead the effort in developing tools and techniques for automatic metadata generation. Funding for her recent research has come from the Texas Advanced Research Program and Google (specifically for a project relevant to this project called *Finding important information in unstructured text*).

In addition to the investigators and researcher, the following UNT staff will participate on a part-time basis in the proposed work.

**Dr. Philip Turner**, Vice Provost for Learning Enhancement, serves on the UNT academic leadership team. He has twenty years of senior leadership experience in higher education with seventeen years

leading major digital learning projects. He currently serves as the Director of UNT's Quality Enhancement Plan which involves a seven-year course redesign effort. Dr. Turner will provide expert advice in the areas of intellectual property, licensing, and related policy areas. He was instrumental in developing intellectual property policies at UNT for faculty members creating online courseware.

**Dr. Ron Carriveau**, Assessment and Measurement Specialist. Dr. Carriveau has extensive experience in all aspects of evaluation of learning outcomes. He is currently working on the Stage 1 project with responsibilities for developing and implementing use scenario to assess use-level functionality of repository. In the proposed project, Dr. Carriveau will provide assistance in ensuring the LOR training materials developed address learning outcomes and also will review methodologies for usability testing.

Much of the project work will involve UNT graduate students from TxCDK, the School of Library and Information Sciences, and the Department of Computer Science and Engineering. These students will assist in technical development and enhancements of the LOR, programming, data collection and analysis related to user requirements and usability testing, developing policies and procedures, migrating/implementing the LOR at the TDL, developing training materials, and other project activities. The will be responsible for populating the LOR with content and creating high-quality metadata records to describe the learning objects. These students have experiences in digital repository implementation, programming, research methods, and metadata.

In addition, external consultants will participate in two aspects of the project: the development of policies for intellectual property and licensing; and training project staff on DSpace Manikin and the implementation of the near-production ready LOR at the TDL.

The project team collectively has the knowledge, skills, and experience to successfully accomplish project goal and objectives. The staff size is commensurate with the work outlined in this proposal and the intended project deliverables. The interdisciplinary team for this project encompasses the domains of Library and Information Sciences and Computer Science, policy and learning assessment, which yields a foundation grounded in competence and innovation.

TxCDK and UNT are institutionally ready to carry out the proposed project. TxCDK and UNT will provide the necessary computing infrastructure for the technology components of the project. TxCDK will perform the project management, including detailed project planning, oversight, and monitoring all project activities leading to the successful completion of the project. UNT has the largest online enrollment of any four-year public institution in Texas with 31 programs online and over 12,000 students enrolled in online courses. The institution has in place the technological infrastructure, intellectual property policies, funding incentives, training programs, instructional design and production support to enable its faculty to utilize emerging information technologies to enhance learning.

TxCDK, in addition to its work on the Stage 1 proof-of-concept LOR, is also piloting two repository applications: an institutional repository application called STARchive <<http://meta.lis.unt.edu/stararchive>>, and a metadata education and research information commons called MERIC <<http://meric.lis.unt.edu>>. Both of these use DSpace as the repository platform. This combined with expertise and experience with metadata-driven web-based applications, and current metadata standards and practices provide the knowledge and skill base for carrying out this project. The project team collectively has the expertise and experience to address the project objectives including automatic metadata generation, usability testing, evaluation, and developing training materials. External consultants will be brought into the project to facilitate the development of intellectual property and licensing policies, and training project staff on DSpace Manikin and the migration/implementation of the LOR at the Texas Digital Library.

## 11. Proposal Summary

This proposal describes a project to lay a solid technical, organizational, and policy foundation for a near-production ready Texas Course Redesign Learning Object Repository. This project builds on the currently funded proof-of-concept learning object project now underway. The project will address the following objectives:

1. Develop and implement a near-production ready LOR application reflecting enhancements based on Stage 1 outcome and insights.
2. Design and customize the LOR user interfaces (end user interfaces and content submitter interfaces) and assess those interfaces using accepted practices and criteria for usability.
3. Examine and implement where appropriate computational methods and processes to assist in automatic metadata generation.
4. Develop, in consultation with THECB staff, draft operational and administrative procedures for a production-level LOR system.
5. Explore opportunities for the Texas Digital Library (TDL) to serve as host for the LOR by implementing the near-production ready LOR on systems operated by the TDL.
6. Initiate, coordinate, and lead discussions, in consultation with THECB staff, with potential stakeholders of the Texas Course Redesign Learning Object Repository, related to intellectual property rights, licensing, and related policies, resulting in a draft of such policies for consideration of THECB and the stakeholders.
7. Ingest learning objects into near-production ready LOR drawing on course content from 4-6 of the redesigned courses (as available for the proposed project).
8. Develop an appropriate training program and/or training materials to assist end users and content submitters in using the LOR.
9. Develop a draft plan, in consultation with THECB, to market the LOR to intended users and other potential customers.
10. Assess the achievement of project goal and objectives, and develop a plan for ongoing evaluation of the LOR when it is in production.

The proposed work is multi-faceted consisting of several, complementary work areas: Project Management, Documentation, and Communications; Technical Development and Enhancements; Users and Usability; Implementation and Operation; Policies and Procedures; Training and Marketing; and Assessment and Evaluation. The project, when completed, will deliver the following:

- The implementation at the Texas Digital Library of a tested and near-production ready LOR application populated with learning objects from 4-6 redesigned courses
- Natural language processing tools for automatic metadata generation and insertion of resulting metadata into LOR metadata records
- Written policies and procedures for the operation and administration of the LOR
- Written policies related to intellectual property, licensing and related topics that reflect the interests of THECB and other key stakeholders
- Appropriate training materials for users of the LOR
- A marketing plan for the LOR
- An evaluation of the success of the project in achieving its goal and objective
- An outcomes-based evaluation plan for ongoing assessment and evaluation of the LOR once it is in full production
- Interim and final reports required by THECB.

The proposed work will be carried out by an interdisciplinary team located at UNT and supplemented by expert consultants. Building on the successful Stage 1 proof-of-concept LOR, we will assist THECB in leveraging its investment in the redesigned courses by making those resources available in a learning object repository designed for the intended users and governed by appropriate policies and procedures.

## References

Nielsen, Jakob. (1993). Usability engineering. Boston: Academic Press.

Rubin, Jeffery. (1994). Handbook of Usability Testing. New York: Wiley, 1994.

Texas A&M Libraries. Digital Initiatives. (2007). DSpace Manakin <<http://di.tamu.edu/projects/xmlui>>

Texas Digital Library. 2007. About the TDL. <<http://www.tdl.org/about/>>

## **Appendix A: Letter of Support from Texas Digital Library**

[Note: A hardcopy of this letter was sent directly to THECB staff.]



June 22, 2007

Kevin Lemoine  
Senior Program Director  
Division of Academic Affairs and Research  
Texas Higher Education Coordinating Board  
P.O. Box 12788  
Austin, TX 78711

Dear Dr. Lemoine:

As Co-Directors of the Texas Digital Library, we are writing in support of the proposal, *The Texas Course Redesign Learning Object Repository: Research and Development for a Production System*, submitted by the Texas Center for Digital Knowledge (TxCDK) at the University of North Texas. Specifically, we want to confirm that the description in the proposal that involves the Texas Digital Library (TDL) is accurate and acceptable. TDL provides a digital infrastructure for the scholarly activities of Texas universities, and this infrastructure includes not only digital repositories but also a preservation network to ensure long-term stewardship of valuable and vital digital resources.

Dr. William E. Moen, the principal investigator for the proposed project, serves as the chair of the TDL Metadata Working Group. He notified us when he received the award for the proof-of-concept learning object repository (LOR) project, and he indicated to us then a potential intersection between such a LOR and TDL's vision for operating a LOR for institutions of higher education in Texas. We are pleased that the Texas Higher Education Coordinating Board (THECB) is considering a LOR for its redesigned courses. We believe that TxCDK's foundational work for the THECB LOR can inform the development and implementation of the TDL LOR, especially in the areas of metadata, policies, and operation.

In support of Dr. Moen's proposal, TDL is offering to host the next version of the THECB LOR resulting from the proposed project. Dr. Moen's choice of DSpace as a repository platform conforms to our current repository infrastructure. We also support the research and development activities in the project that relate to DSpace Manakin, the XML-based user interface application developed at Texas A&M (the home of Dr. Leggett, TDL Co-Director). TDL plans to use DSpace Manakin in its repository applications to enable easier and more useful customization for the various digital collections we envision being components of TDL.

While the details of hosting the THECB LOR will be worked out if the TxCDK proposal is approved, we plan to provide the following in support of this proposal:

- A DSpace implementation in a test environment (i.e., not fully publicly available) to host the version of the THECB LOR resulting from the proposed work.
- Implementation of the operational policies developed in the proposed project (e.g., access controls, submission policies, metadata scheme, etc.)
- Appropriate access to the THECB LOR as required by THECB, the project, etc., for testing and assessment
- Consultation in regards to the use of Manakin and other DSpace technologies.

We believe that the proposed project can result in critical synergies to benefit Texas higher education. The deliverables from the proposed project will leverage THECB's investment in the redesigned courses by moving forward towards a production-level LOR. TDL will benefit from the experience of implementing

the LOR. When THECB makes the decision to provide an operational LOR, TDL will be available to host and operate the system and provide the preservation network to support long-term stewardship for these valuable learning objects.

Sincerely,



Mark McFarland  
John Leggett  
Co-Directors, Texas Digital Library

## Appendix B: Automating the Generation of Metadata for Learning Objects

This appendix provides an example of how natural language processing (NLP) assists in metadata generation. What follows is first a selected “document” (that contains text from a learning object (LO)). Several automatic processes act on the document. The first output is a set of **keywords** extracted based on a NLP algorithm that identifies and extracts words and phrases that indicate the topic of the LO. Next, operating on the document and the keywords, another algorithm takes specific **concepts** represented by the keywords and puts them in a larger context of related concepts resulting in an **ontology**, a knowledge structure that can assist users in navigating LOs to identify relevant materials. The third operation is automatic summarization of the document. NLP algorithms generate a summary of the document. The minimizes the need for human-generated abstracts and summaries. Such summaries can assist the user to read a brief overview of a particular LO.

### Document

The third component of world war was the largest and most sanguinary of all. Hitler's appetite for imperial conquest had always been directed eastwards to the USSR with its vast supplies of food, materials, manpower, and territory to colonize. In December 1940 Hitler turned away from Britain and approved BARBAROSSA, the large-scale invasion of the USSR. The motives for the contest were not only imperial. Soviet communism represented a profound social and political threat and Hitler, an ardent anti-communist throughout the inter-war years, saw the final contest with Marxism as a necessity. Following the German-Soviet pact of 1939 the threat became greater. A Soviet-Finnish war in the winter of 1939-40 resulted in Soviet encroachments in the Baltic. In June 1940 Soviet troops occupied the Baltic states and seized Bessarabia from Romania. Hitler ordered the conquest of the USSR before it became too entrenched in eastern Europe.

### Keywords

Barbarossa, World War, Hitler, USSR invasion, German-Soviet pact, Soviet-Finnish War, Bessarabia.

### Concepts and ontology

Barbarossa, USSR invasion  
German-Soviet pact  
Soviet-Finnish war  
    → Events in World War II  
    → Events in History

Hitler

- German chancellor
- Leader
- Founder of Nazism

Bessarabia

- Region in Romania

### Summary

Hitler turned away from Britain, and approved the invasion of USSR. He considered social communism as a social and political threat.

## Appendix C: Brief Curriculum Vitae of Investigators and Researcher

This appendix contains brief curriculum vitae for the three investigators:

- **Principal Investigator:** William E. Moen, Ph.D. Texas Center for Digital Knowledge, University of North Texas
- **Co-Principal Investigator:** Ms. Vandana Singh (ABD) Texas Center for Digital Knowledge, University of North Texas
- **Natural Language Processing Expert:** Dr. Rada Mihalcea, Department of Computer Science, University of North Texas

**William E. Moen, Ph.D.**

School of Library and Information Sciences  
Texas Center for Digital Knowledge  
University of North Texas  
P.O. Box 311068, Denton, TX 76203-1068  
email: wemoen@unt.edu website: <http://www.unt.edu/women>  
voice: 940-565-3563 fax: 940-565-3101

Complete CV is available at: <http://www.unt.edu/wmoen/VitaCurrent.pdf>

**ACADEMIC PREPARATION**

<b>Institution</b>	<b>Major</b>	<b>Degree</b>	<b>Date</b>
Syracuse University	Information Transfer	Ph.D.	1998
Louisiana State University	Library and Information Science	M.L.S.	1988
University of Montana	Philosophy	B.A.	1987

---

**TEACHING AND PROFESSIONAL EXPERIENCE**

**Associate Professor.** School of Library and Information Sciences. University of North Texas. 2002–  
Areas of teaching and research include the organization of information; metadata and networked  
information organization and retrieval; Internet applications and services; Z39.50.

**Fellow.** Texas Center for Digital Knowledge. University of North Texas. 2001–.

**Assistant Professor.** School of Library and Information Sciences. University of North Texas. 1998–2002.

**Instructor.** School of Library and Information Sciences. University of North Texas. 1996–98.

**Network System Research Analyst.** Network Development and MARC Standards Office, Library of  
Congress. 1989-1991.

---

**PUBLICATIONS (Selected)**

**Journal Issues (Special Issue Editor)**

**Metadata: A Networked Information Strategy to Improve Access to and Management of  
Government Information** [Symposium Issue]. *Government Information Quarterly*, 18(3) July 2001.

**Chapters in Books/Journal Articles**

**The Challenges of Nonstandardized Vendor Usage Data in a Statewide Metasearch Environment:  
The Library of Texas Experience** (with Fatih Oguz and Charles R. McClure). *Library Quarterly*, v  
74, no 4 October 2004: 403-422.

**A Service-Based Approach for Virtual Libraries** (with Kathleen R. Murray). *Texas Library Journal*. Fall  
2002. Available URL:

**Designing and Demonstrating a Resource Discovery Service for the Library of Texas** (with Kathleen  
R. Murray). *Texas Library Journal*. Fall 2002.

**The Metadata Approach to Accessing Government Information.** *Government Information Quarterly*,  
18(3) July 2001: 155-165.

**Assessing Interoperability in the Networked Environment: Standards, Evaluation, and Testbeds in  
the Context of Z39.50.** In *Evaluating Networked Information Services: Techniques, Policy, and  
Issues* (pp.85-109), edited by Charles R. McClure and John Carlo Bertot, Publisher: ASIS thru  
Information Today, Inc. 2001.

**Interoperability for Information Access: Technical Standards and Policy Considerations.** *Journal of  
Academic Librarianship*, 26(2) March 2000: 129-132.

**Published Conference Papers**

**An Extensible Approach to Interoperability Testing: The Use of Special Diagnostic Records in the  
Context of Z39.50 and Online Library Catalogs** (with Sebastian Hammer, Mike Taylor, Jason  
Thomale, and JungWon Yoon). In *Proceedings of the 68<sup>th</sup> ASIS&T Annual Meeting*, v 42. Charlotte,  
North Carolina, October 28 – November 2, 2005. Silver Spring, MD: American Society for  
Information Science and Technology.

**Design, Development, and Implementation of a Texas Library Directory Database: A Multipurpose Database for the Library of Texas** (with Irene Lopatovska and Fatih Oguz). In *Proceedings of the 67<sup>th</sup> ASIS&T Annual Meeting*, v 41. Providence, Rhode Island, November 12-17, 2004 (pp. 93-1001). Medford, NJ: Information Today, Inc. 1997.

**Optimizing Resource Discovery Service Interfaces in Statewide Virtual Libraries: The Library of Texas Challenge** (with Kathleen R. Murray and Irene Lopatovska). In *Proceedings of Libraries Without Walls 5: The Distributed Delivery of Library and Information Services*. Molyvos, Greece, September 19-23, 2003. London: Facet Publishing. 2004.

**Assessing Metadata Utilization: An Analysis of MARC Content Designation Use** (with Penelope Benardino). In *2003 Dublin Core Conference: Supporting Communities of Discourse and Practice – Metadata Research and Application*. Seattle, WA, September 28-October 2, 2003. Seattle: Information School of the University of Washington, 2003.

**Mapping the Interoperability Landscape for Networked Information Retrieval**. In *Proceedings of First ACM/IEEE-CS Joint Conference on Digital Libraries*. Roanoke, VA, June 24-28, 2001 (pp. 50-52). New York: The Association for Computing Machinery, 2001.

---

### **RESEARCH ACTIVITIES (Selected)**

**Co-Principal Investigator** (with Dr. Jiangping Chen). Authors @ Your Library: Web Database System.. Funded by the Association of American Publishers. A 12-month research and development project to create a match-making service for library programs and publishers and their authors. 2005-2006.

**Principal Investigator**. Examining Present Practices to Inform Future Metadata Use: An Empirical Analysis of MARC Content Designation Utilization. Funded by a National Leadership Grant from the U.S. Federal Institute of Museum and Library Services. A 21-month project to investigate library catalogers' use of MARC content designation. 2004-2006. Extended to August 2007.

**Principal Investigator**. Developing an Alternative Approach for Interoperability Testing of Library Z39.50 Servers. Funded by the U.S. Federal Institute of Museum and Library Services as an extension to the National Leadership Grant, Realizing the Vision of Networked Access to Library Resources. An 8-month project to develop an alternative method of interoperability testing using specially-created MARC diagnostic records. 2004-2005.

**Principal Investigator**. System Configuration for Implementing the LOT Resource Discovery Service. Funded by the Texas State Library and Archives Commission. A project to begin implementation of the LOT resource discovery service for the Library of Texas. 2003.

**Principal Investigator**. Needs Assessment Study of Texas Academic, Public, and School Libraries: Funded by the Telecommunication Infrastructure Fund (TIF) Board. A 9-month project to survey Texas libraries to inform the TIF Board and its Library Working Group on the state of these libraries in terms of technology, skills, and other needs. 2002.

**Principal Investigator**. The Z Implementation Component of the Library of Texas. Funded by the Texas State Library and Archives Commission. A 10-month project to develop the virtual catalog and integrating search/retrieval interface for the Library of Texas. 2001-2002.

**Principal Investigator**. Illinois Library Resources and Library Automation in the Networked Environment: An Analysis and Recommended Strategies. Funded by the Illinois State Library. A 7-month project to assess shared library automation and recommend strategies for statewide access to library resources. 2001-2002.

**Principal Investigator**. Realizing the Vision of Networked Access to Library Resources: An Applied Research and Demonstration Project to Establish and Operate a Z39.50 Interoperability Testbed. Funded by the U.S. Federal Institute of Museum and Library Services. A 20-month project to improve interoperability in search and retrieval with online library catalogs. 1999-2001. Extended through December 2003.

---

### **SELECTED SCHOLARLY & PROFESSIONAL ACTIVITIES & ACHIEVEMENTS**

**Frederick G. Kilgour Award for Research in Library and Information Technology**. Library and Information Technology Association (LITA) and OCLC Online Computer Library Center. 2005.

**Interim Director**. Texas Center for Digital Knowledge, University of North Texas. 2004-present.

**Fellow**. Texas Center for Digital Knowledge, University of North Texas. 2001-present.

**Chair. Metadata Working Group for the Texas Digital Library**, 2007-present.

**Vandana Singh**

School of Library and Information Sciences, Texas Center for Digital Knowledge  
University of North Texas  
P.O. Box 311068, Denton, TX 76203-1068  
[vandana.singh@gmail.com](mailto:vandana.singh@gmail.com)

**Education**

---

**University of Illinois at Urbana-Champaign**

Ph.D. Information Science Aug 2007 (*expected*)

**University of Chicago**

M.S. Computer Science June 2006

**Wageningen University, The Netherlands**

M.S. Knowledge Management Systems March 2001

**Pantnagar University, College of Ag., India**

B.S. Economics (major) June 1998

**Research Experience**

---

**Research Assistant, University of Illinois at Urbana-Champaign**

**1. Technical Support for Software**

This work was supported by a National Science Foundation funded project, "Over The Shoulder Learning". The aim of this project was to study the problem solving techniques used by people in collaboratively solving software issues. This study helps in defining the design of user friendly systems. I conducted ethnographic research on three different levels of technical support under this project:

- *Face to face*  
In the face to face research project we focused on understanding the work place interaction between employees, trying to solve problems in using CAD software. We studied effective knowledge sharing activities of the employees and proposed requisite changes in the software that could ease the learning curve of new employees. This study was conducted at Caterpillar Inc and State Farm.
- *Over telephone*  
We conducted ethnographic study of a unit of State Farm Corporation that dealt with handling telephone calls to solve the problems of their company wide intranet. Their agents in different parts of the country use their intranet to track technical problems in the offices of their agents and do the work of scheduling the technical support employees to these locations. This study was conducted at Caterpillar Inc and State Farm.
- *Online (on Internet in Open Source Software)*  
This study was aimed towards understanding the same process of technical support provided to users but it was in an online context. The availability of the discussion lists and support forums on the websites of Open Source Software made it easy to analyze the process of help giving in digital medium.

**2. Knowledge Management Systems**

- *In State Farm*  
Collected data on in-house users of a knowledge base for the usability group at State Farm, Bloomington, and presented the results of data analysis. We studied the adoption of the knowledge base, the problems that the employees were having in understanding this system and the methods they were adopting to solve these problems.

- *At NCSA*  
Worked at National Center for Supercomputing Applications (NCSA) with the Knowledge and Learning Systems group. The group has an objective of improving information systems by developing better visualization techniques and identifying parameters for evaluation of information quality for knowledge management and e-learning systems.
- 3. Collaborative software**
    - *At Microsoft*  
Microsoft provided us with the beta version of their software Share Point Team Server. We worked in teams to evaluate the usability of the system, explain the shortcomings, propose recommendations and develop prototypes for the recommendations. We used cognitive walkthrough, heuristics evaluation techniques and scenario based design for this study. I presented the results of this study at Microsoft, Redmond November 2002.
  - 4. Usability Research Experience**
    - In a course on interfaces I designed a user study, conducted usability test, worked towards iterative web design by using rapid prototyping techniques to present an alternative website for the then existing engineering library interface at University of Illinois Urbana Champaign.
  - 5. Community Informatics**
    - I made a database of relevant conferences, key journals, key articles, relevant national and international organizations and key people in the field of Community Informatics. This was done to establish a common point for these resources and also to invite and involve the main people in Community Informatics community to the Initiative. This was used as the starting point in the current Community Informatics Initiative at University of Illinois Urbana-Champaign.

## Publications

---

1. Rathi, D., Twidale, M. B., **Singh, V.**, and Singh, A. B. "Lessons Learned and Shared: A case study of the development and deployment of a Knowledge Management System" (*Under Review* Submitted to International Journal of Business Information Systems).
2. **Singh, V.** "Open Help for Open Source: how a community shares technical help online" Doctoral Student Research Poster Competition, ALISE 2007, Seattle, Washington, January 15-19, 2007.
3. Rathi, D., Twidale, M. B. **Singh, V.** and Jones, M. C. "Your Mark is My Dirt: Impact of Email Signatures on Decision Making", 9th International Conference on Decision Support Systems, January 2-4, 2007 , India.
4. **Singh, V.**, Twidale, M., Rathi, D., "Open Source Technical Support – A look at Peer Help Giving". Hawaii International Conference on System Science (HICSS – 39), IEEE, January 2006.
5. Rathi, D., Twidale, M., **Singh, V.** "Adapting Knowledge Management Systems for Collaborative Technical Problem Solving". International Conference on Knowledge Management (ICKM), ASIST, October 2005.
6. Rathi, D., **Singh, V.**, Gebauer, J., "Ubiquitous Computing for Knowledge Management." Knowledge Sharing and Collaborative Engineering (KSCE 2004), IASTED, November 2004.
7. Jones, M. C., Medina, K. E., Rao, A., Rathi, D., **Singh, V.**, "I-Vote: an audience voting system." Student competition paper. CHI extended abstracts on human factors in computing systems, April 2004
8. Technical Report by TEC lab, Speech Communications.  
Contractor, N., Monge, P., Fulk, J. "Knowledge Networks and Distributed Intelligence" Brandon D., Dandi R., Huang, M., Palazzolo, E., **Singh, V.**, Su, C. Presented to the Professional Development Workshop Academy of Management Seattle, Washington, August 2003.

**Rada Mihalcea**

Department of Computer Science and Engineering  
University of North Texas  
Denton, TX 76203

email: rada@cs.unt.edu      website: <http://www.cs.unt.edu/~rada/>      voice: 940-369-7630

**Education**

- Ph.D., Computer Science and Engineering, Southern Methodist University, 2001.
- M.S., Computer Science and Engineering, Southern Methodist University, 1999.
- B.S., Computer Engineering, Technical University of Cluj-Napoca, Romania, 1997.

**Professional Appointments**

- University of North Texas, Computer Science Department, Assistant Professor, 2002 - present.
- MIT Media Arts and Sciences, Visiting Scholar, summer 2005.
- University of Texas at Dallas, Computer Science Department, Visiting Professor, spring 2002.

**Publications Related to Proposed Project**

- Andras Csomai and Rada Mihalcea, *Investigations in Unsupervised Back-of-the-Book Indexing*, in Proceedings of the Florida Artificial Intelligence Society, 2007.
- Andras Csomai and Rada Mihalcea, *Linking Educational Materials to Encyclopedic Knowledge*, in Proceedings of the Conference on Artificial Intelligence in Education, 2007.
- Rada Mihalcea and Hakan Ceylan, *Explorations in Automatic Book Summarization*, in Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP 2004, Prague, Czeck Republic, June 2007.
- Rada Mihalcea and Paul Tarau, *TextRank: Bringing Order into Texts*, in Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP 2004, Barcelona, Spain, July 2004.
- Rada Mihalcea and Samer Hassan, *Using the Essence of Texts to Improve Document Classification*, in Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP), Borovetz, Bulgaria, September 2005.

**Other Relevant Publications**

- Rada Mihalcea, Courtney Corley, Carlo Strapparava, *Corpus-based and Knowledge-based Measures of Text Semantic Similarity*, in Proceedings of the American Association for Artificial Intelligence (AAAI 2006), Boston, July 2006.
- Rada Mihalcea, *Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization*, in Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics ACL 2004, Barcelona, Spain, July 2004.
- Rada Mihalcea, *Word Sense Disambiguation Using Pattern Learning and Automatic Feature Selection*, Journal of Natural Language and Engineering, December 2002.
- Rada Mihalcea, *Going Beyond Explicit Knowledge for Improved Semantic Based Information Retrieval*, International Journal on Artificial Intelligence Tools, December 2002.
- Dan Moldovan and Rada Mihalcea, *Improving the Search on the Internet by Using WordNet and New Operators*, IEEE Internet Computing, January 2000.

**Relevant Research Experience**

- Finding important information in unstructured text. Funded by Google, Inc.
- *TextRank*. Graph-based ranking algorithms for text processing. Funded by the Texas Advanced Research Program.
- *Babylon*. Building and exploiting parallel texts for languages with scarce resources. Collaboration with the National Polytechnic Institute in Mexico. Funded by the North Texas Hispanic and Global Studies Fund.

### Synergistic Activities

- ACL-SIGLEX President, 2004-present.
- Senseval-3 coordinator (with Phil Edmonds).
- ACL-SIGNLL board member, 2002-present.
- Editorial boards: Computational Linguistics (2005-2008), Language Resources and Evaluations (2005-2007), Journal of Natural Language Engineering (2007-2011), Research in Language and Computation (2007-2011).
- Co-chair *Building and Exploiting Parallel Texts: Data Driven Machine Translation and Beyond*, ACL 2005 workshop; HLT/NAACL 2003 workshop.
- Selected recent program committees: ACL 2007, IJCAI 2007, CICLING 2007, AAAI 2007, NAACL 2007, ACL 2006, EMNLP 2006 (area chair), NAACL 2006, CICLING 2006, AAAI 2006, CONLL 2006, EMNLP 2005, ACL 2005, CONLL 2005, IJCAI 2005, RANLP 2005, CICLING 2005
- Reviewer for Computational Linguistics, ACM, CLS, IJPRAI, IJAIT, JAIR, JNLE

### Collaborators

[Current Students:] Carmen Banea (Ph.D.), Andras Csomai (Ph.D.), Hakan Ceylan (Ph.D.), Kino Coursey (Ph.D.), Andras Csomai (Ph.D.), Samer Hassan (Ph.D.), Balathasan Giritharan (M.S.), Christian Loza (Ph.D.), Michael Mohler (M.S.), Ravi Sinha (M.S.)

[Graduated Students:] Yohan Chandra (M.S.), Ben Leong (M.S.), Ehsan Faruque (M.S.), Lei Shi (M.S.), Li Yang (M.S.).

[Collaborators past 48 months:] Nathaniel Ayewah, Timothy Chklovski, Pedro Domingos, Phil Edmonds, Elizabeth Figa, Roxana Girju, Sanda Harabagiu, Adam Killgariff, Philipp Koehn, Henry Lieberman, Hugo Liu, Joel Martin, Dan Moldovan, Christof Monz, Paul Morarescu, Vivi Nastase, Marius Pasca, Ted Pedersen, Amruta Purandare, Drago Radev, Vasile Rus, Michel Simard, Push Singh, Carlo Strapparava, Mihai Surdeanu, Paul Tarau, Doina Tatar, Jan Wiebe].

## Appendix D: Budget