
“A Testbed of Civil War-Era Newspapers”

IMLS grant #LG-02-03-0082-03

Semi-annual report, October 2005-March 2006

Submitted by: James Rettig, PI
University Librarian
Boatwright Memorial Library
University of Richmond

Introduction

The University of Richmond and its partner the Perseus Project of Tufts University continue to make progress on this project. This report addresses the following highlights of the past half year:

1. Status of the digital librarian position
2. Installation and Data Ingest in XPAT/DLXS
3. Status of Richmond *Daily Dispatch* Content Files
4. Status of *The Liberator* Files
5. Status of *The Liberator* Files
6. Status of *Philadelphia Ledger* Files
7. WebWise 2006
8. Work on outcomes assessment using the project's Logic Model
9. Named entity tagging
 - Overview
 - Publications
 - Evaluation of tagging
 - Named entity system progress
 - Authority lists and supplementary sources

1. Status of the Digital Librarian Position

This IMLS grant provided funding for a Digital Resources Librarian position at the University of Richmond. We hired Dr. Andrew Rouner at the beginning of this grant and his contributions to our project are immeasurable. Through his talents and expertise we learned about XML, TEI encoding, repository systems, indexing systems, and digitization workflow processes including versioning tools and standard tagging protocols. Andrew resigned from this position in February, 2006 to take on the opportunity as Director of Digital Libraries at the Washington University in St. Louis. Although he is in Missouri, he has volunteered his time as a consultant until his position is filled. As part of its commitment to sustain the work and momentum of this grant, the University of Richmond has committed to a full time permanent digital librarian position. This position has been approved by Human Resources and a national search is underway. We have interviewed several strong candidates and hope to have a person hired by May 15th, if not earlier. As we complete the newspaper project, we will use the remaining salary monies in the grant towards the salary of the person hired into this new permanent position.

During the interim, Rick Neal, Richmond's library systems manager, has stepped in to help keep us up and running. He has done an outstanding job. Rick has become proficient with the tools needed to load and transfer large amounts of data. One of our student employees has become knowledgeable enough about TEI encoding that we have created an expanded role for him to contribute to our project. This role will actually be a summer intern position that we hope will qualify for credit towards his master's degree in library science. He will start library school in the fall at the University of North Carolina at Chapel Hill.

2. Installation and Data Ingest in XPAT/DLXS

The decision to license XPAT/DLXS was made by the University of Richmond (UR) in the spring of 2005, and a license was purchased at the end of the university fiscal year in early summer. This decision was made primarily for two reasons. First, since the Richmond *Daily Dispatch* files were encoded at a high level in TEI XML, and were then set to receive additional metadata on personal and place-names from the Perseus Project, only very robust indexing software could leverage this rich metadata. Second, since the project was undertaken under the auspices of an IMLS grant, completion of the project within the stated timeframe was a priority, and it was decided that working with supported and relatively mature software was the prudent course.

DLXS annually offers a five day workshop on its software, usually in August, at the University of Michigan campus in Ann Arbor. It is open to licensees who pay the annual subscription fee. Its current version of the software at the time of licensing was 11a, but DLXS promised to deliver version 12 prior to the August workshop. In fact, the workshop dealt exclusively with version 12. DLXS/XPAT 12 was needed over against version 11 for our data for two primary reasons. First, Release 12 supported non-continuous tone images, including the jpeg2000 format (which allows users to pan and zoom images), and version 12 also supported not only XML, but Unicode entities and UTF-8 text. From the outset of the project, UR had adopted Unicode XML entities and UTF-8 text as the latest standards in encoding texts. The compressed image format would allow the University of Richmond to fit all the images associated with this project on our available server space, whereas UR did not have similar space available for the original tif images (which was the format supported by DLXS in Release 11). Release 12 was therefore essential for the project.

Unfortunately, DLXS 12 was not released until October 31, 2005. Installation of XPAT/DLXS took place over the two weeks. (We previously had installed DLXS 11a, and this did help to streamline the Release 12 installation.) This took longer than it might otherwise have, because the software was new and there were problems with it as delivered (one problem we encountered was made noted by DLXS for other users in this phase). For the remainder of November and through most of December (up to the break) we attempted to load a sample of our data—10 newspaper issues and associated images from the Richmond *Daily Dispatch*. We were ultimately successful in this, but exchanged probably over one hundred emails with DLXS support in getting this to work. Again, at several points, modifications to the code were required to make the software run as hoped. Once the sample subset loaded properly, one attempt was made to load all the issues of the *Daily Dispatch*, which was ultimately not successful. At that point, Dr. Andrew Rouner, who had been working with DLXS in a grant-funded position at UR, left to accept another position at Washington University in St. Louis. His position should be filled at some point in May or June, and work on DLXS can continue at that point.

The process for loading data into DLXS/XPAT can be fairly involved. Although it is geared toward texts encoded in TEI, such encoding varies from project to project, and DLXS requires that XML files conform to its file-naming and directory structure conventions, as well as file format standards and DTD. UR semi-automated this process, including batch-renaming of files, batch-processing of minor changes internal to the files via PERL scripts (which involved changing, i.e., <milestone> tags that marked page-breaks, to <pb> tags, capitalization of all tags, changing certain <id> tags, etc.) It further required regularization of tags (via the OSX program) and conversion of the text to UTF-8. Most of these changes require a change in DTD as the process moves ahead, but the files also needed to be validated in the course of the process, which made the process that much more complicated. Ultimately, making all these changes with roughly 1400 files proved more difficult than making them with 10. That said, the process for transforming this set of files has been documented and semi-automated, so once the position is filled, adaptation of that process should be relatively smooth.

3. Status of Richmond *Daily Dispatch* Content Files

All of the Richmond *Daily Dispatch* has been checked by the University of Richmond and were made available on March 31, 2006 to Tufts for automatic named entity encoding. Tufts will be loading these enriched files into Perseus and send copies back to Richmond for loading into DLXS.

4. Status of *The Liberator* Files

The Liberator consists of archival TIF files, JPEG 2000 deliverables, and flat text files. Richmond is processing the .txt files, adding TEI headers, converting to XML. These files will then be sent to Tufts for basic TEI encoding, plus the name entity tagging. Upon completion, these files will be delivered back to Richmond for loading into the DLXS repository. We are striving to have this workflow completed by the beginning of July.

5. Status of *Philadelphia Ledger* Files

The Philadelphia Ledger was delivered to Richmond in April 2006. This collection currently consists of archival TIF files and raw OCR'd text files. The University of Richmond will create the deliverable JPEG 2000 files and analyze the text data for the best process to proceed. We anticipate creating a TEI header for these issues, then delivering them to TUFTS for treatment similar to *The Liberator*. We plan to have this workflow completed by end of August.

6. WebWise 2006

The University of Richmond and Tufts University sent representation to the WebWise annual conference in Los Angeles, California. Rachel Frick and Jim Rettig from Richmond and Alison Jones from Tufts attended the conference. During a designated portion of the conference we were able to report on our progress of our project. The University of Richmond showed examples of the main gateway to our project and demonstrated DLXS with the sample collection of newspapers. Tufts also presented an abstract of its research on automatic named entity tagging, as well as demonstrating the *Richmond Daily Dispatch* data as it presented through the Perseus interface. We receive positive feedback on our project content, but more importantly on our research that we are conducting in the context of our project.

Rachel Frick also took the opportunity at WebWise to consult with colleagues at the Colorado and Utah newspaper projects. John Herbert, Utah, and Jill Koelling, Colorado, confirmed that they were maintaining statistics generated from our jointly designed user survey. These results will be compared to data culled from Richmond's site. Utah and Colorado will also provide cost data to Richmond for analysis in the Richmond final report. Rachel Frick also met with Natalia Smith, the Digitization Librarian at the University of North Carolina at Chapel Hill. They discussed possible connections between collections and potential future collaborations. John Herbert also encouraged Tufts/Richmond to submit a proposal for the IFLA Newspaper conference in Utah this coming May. Greg Crane will be submitting a proposal on the project's behalf. WebWise was a great opportunity to unveil the beginnings of our project, as well as connect with colleagues who are creating similar collections, either around format(newspapers) or content (Southern and Civil War history).

A copy of the handout distributed during the demonstration session at WebWise is appended to this report.

7. Work on outcomes assessment using the project's Logic Model

As explained above, delays in release of the new version of DLXS software and the loss of the grant-funded Digital Librarian have caused delay the public debut of the project's Web site offering access to digitized content. This forces a revision to the schedule for some assessment activities. Appendix #2 to this report is a revised version of our assessment logic model. Dates highlighted in lavender are the revised dates.

8. Named Entity Tagging

Overview

The last six months have focused on two major areas: evaluation of the named entity tagging in the newspaper files and writing up the results of both this work and our research in general on the related Perseus American Collection.

Publications

In the last six months, we have completed a number of publications reporting on the work conducted on the named entity system, the Civil War newspapers and the American Text Collection. In the fall, a white paper on the technical details of the named entity system used with the newspapers and an evaluation of its initial results with the Perseus American Collection was finished. This report has been ingested into Tufts digital repository and is available on the Web.¹

A paper reporting on our named entity work with the *Richmond Times Dispatch* has been accepted for publication by the Joint Conference on Digital Libraries (JCDL), and we will present on our work at this conference in June 2006. The paper entitled, "The Challenge of Virginia Banks: An Evaluation of Named Entity Analysis in a 19th-Century Newspaper Collection" evaluates the automatic extraction of ten classes of entities and details the major problems of applying a named entity system to historical texts such as newspapers. This paper also reports on general levels of tagging accuracy which will be discussed in the next section. The paper received very positive reviews and has presented an excellent opportunity to report on the results of our work.

In addition, the Perseus Digital Library was responsible for overseeing the special issue of *D-Lib* in March 2006, which focused on the evolution of digital libraries and includes the Perseus American Collection as the featured collection. Two articles were authored by project staff, "What Do You Do with a Million Books?"² by Gregory Crane, and "Text, Information, Knowledge and the Evolving Record of Humanity"³ by Gregory Crane and Alison Jones. Both of these articles address issues relevant to our work with the Civil War newspapers, including named entity recognition, and the importance of historical knowledge sources for improving the results of automated technologies in creating more intelligent digital libraries.

Evaluation of Tagging

Our last report discussed the different types of entities that we were currently tagging and the major problems associated with each type of entity. Evaluation of this system on the Richmond newspaper and our American collection by project staff continued throughout the fall. A larger evaluation that formed that basis of the JCDL paper was conducted in early January. We explored whether the major sources of error reported last time had been resolved. The work was conducted by project team member Alison Jones and two student workers.

For each of the available 1,355 issues of the *Richmond Times Dispatch*, we calculated the top 20 named entities in each of the ten selected categories, providing a maximum of 200 unique entities to describe each newspaper issue as a whole. Since the newspaper issues are relatively short (about 17,300 words each), many do not mention 20 distinct newspapers or military units and thus generate profiles with fewer than 200 entities. In effect, we provide automatic metadata to illustrate the "aboutness" of each article and issue as a whole. Our focus in this evaluation was thus upon our success with the more frequently cited entities rather than low frequency entities. We were also interested in understanding how much impact limited correction can have upon overall performance. In a number of instances, (particularly with newspapers) we found that the removal of a few obvious and very common errors (such as stopping the tagging of expressions such as *Stars and Stripes*) can dramatically improve accuracy. For our next evaluation, we plan to study the secondary effects of such high impact/low cost corrections on subsequent analysis.

From this automatically generated list of most frequent entities, a random sample of entities was created and the tagging accuracy of each entity was checked. Almost 10,000 entities were manually examined for accuracy in tagging. The following table presents the results of this evaluation:

¹ http://dl.tufts.edu/view_pdf.jsp?urn=tufts:facpubs:gcrane-2006.00001

² <http://www.dlib.org/dlib/march06/crane/03crane.html>

³ <http://www.dlib.org/dlib/march06/jones/03jones.html>

Table: Manual Evaluation of Entity Results

Entity Type	Total Entities Manually Evaluated	Total Correctly Tagged Entities	Total Incorrectly Tagged Entities	Tagging Accuracy	Major Source of Error
Personal Names	2246	1720	526	76.58%	Incorrect Name matching
Places	2592	2525	67	97.42%	Tagging incorrect place, tagging ethnic groups
Dates	1338	1291	47	96.49%	Tagging partial dates
Products	488	281	207	57.58%	Tagging of products/commodities when used as common nouns
Organizations	823	676	147	82.14%	Tagging of general nouns as organizations (offices, army, etc.)
Streets	1210	1202	8	99.34%	Tagging towns such as Brighton, Holyoke as streets
Newspapers	401	267	134	66.58%	Tagging of proper nouns as newspaper titles
Ships	226	211	15	93.36%	Tagging verbs as ship names
Regiments	400	367	33	91.75%	Tagging document sections as military sections
Railroads	126	112	14	88.89%	Tagging of partial railroad names
Total Entities	9850	8652	1198	85.01%	

Our highest level of tagging accuracy is currently with street names due to the addition of a number of successful heuristics (discussed in the previous report), while our lowest level is with the tagging of products (largely due to the generic nature of many commodity names). We found that since our last major evaluation a number of issues have been resolved.

In our last evaluation, we found a number of grammatical patterns that led to incorrect tagging. For example, previously a string of terms such as “manners, morals, court” ending with an organization name type word such as “bank” or “bureau” or “court” would often tag as an organization. This issue has been resolved. Similarly, our tagging of personal names that include honorifics, suffixes and abbreviations is also much improved. Our tagging of commodities with compound names has also improved somewhat due to more robust lists of 19th century commodities but general tagging performance still lags behind others entity categories in tagging accuracy. Place name tagging is also much improved through the elimination of problematic abbreviations and the addition of many local place names such as “Shockoe Slip” to our authority lists. We have also eliminated the tagging of certain types of entities previously defined as “problematic” in our last report, such as “establishments.”

Currently over 1,440,969 entities have been tagged by the automated system, with personal names (542,271), places (401,128) and dates (156,009) being the three most frequently tagged entities. The automatically generated data also shows interesting results: “ammunition,” for example, becomes a prominent commodity in late June 1861, just before serious hostilities begin. Similarly, Robert E. Lee first appears as a significant figure in April 1861 and Gettysburg appears quite frequently at the start of July 1863.

Named Entity System Progress

The named entity system is nearing completion and it is our plan to make it freely available to other interested parties through the Perseus Digital Library website. Named entity tagging on the *Richmond Dispatch* is complete, and the system will next be applied to the *Liberator* and the *Public Ledger* of Philadelphia.

All of the XML files for the *Richmond Dispatch* have been loaded into the Perseus Digital Library System for testing and evaluation, and can be viewed online.⁴ This test system was on display at WebWise. The *Dispatch* can be searched and browsed through our beta named entity browser⁵, which currently allows for searching by personal names, place names and dates. This named entity browser, but not the newspaper collection, was rolled out as part of the new release of our American Text Collection on March 15, 2006.

While there are still more errors than be corrected manually by project staff, the browsing environment currently set up in the Perseus Digital Library will allow for easier error identification than previously. Previous identification of errors involved viewing of XML files in the Oxygen XML editor, which could be confusing at times for students. We are also exploring the creation of an environment (such as a wiki) that will allow users to make corrections to tagging errors as we move forward.

Authority Lists and Supplementary Sources

During the course of this project, a number of “authority lists” have been created to supplement the historical authority lists that we already use to support our named entity system such as George P. Rowell & Co.’s *American Newspaper Directory* (1869) and Henry Varnum Poor’s *History of the Railroads and Canals of the United States* (1860). We have also discovered “authority lists” within other works, such as lists of commodities within *Knight’s Mechanical Encyclopedia* and a list of Civil War era ships from Admiral David D. Porter’s *Naval History of the Civil War*.

The entities that are automatically extracted from the newspaper XML files are currently being stored in a MySQL database, which drives the named entity browser. We are still exploring how metadata formats that are native XML such as MADS records might be used to store and structure this data in the form of a more traditional authority list instead of a relational database. The lists of entities we have extracted could serve other projects well if they were to implement automated tagging, either through the use of our system or through the use of a tool such as GATE (General Architecture for Text Engineering).⁶ In addition, we are exploring how to best transfer the entity data that we have extracted to Richmond to be loaded into the DLXS system.

A number of additional supplementary sources have also been acquired by the Perseus Project in the last few months to supplement the newspaper collection and will ultimately be digitized and transferred to Richmond. These works include *Journal of Alfred Ely: A Prisoner of War in Richmond* (New York: D. Appleton & Company, 1862), *Prison Life in the Tobacco Warehouse at Richmond* (Philadelphia: George W. Childs: 1862), and *A Belle of the Fifties: Memoirs of Mrs. Clay ... in Washington and the South, 1853-1866* (New York: Doubleday, Page & Company, 1905). These are in addition to a number of other sources from our original list of supplementary sources that already have been digitized.

⁴ <http://alkestis.perseus.tufts.edu/hopper/collection.jsp?collection=Perseus:collection:RichTimes>

⁵ <http://www.perseus.tufts.edu/hopper/nebrowser.jsp>

⁶ <http://gate.ac.uk/>

Appendix #1: WebWise 2006 Handout

Civil War Newspapers: Lessons in Digitization and Metadata

On September 24th, 2003 the Institute of Museum and Library Services (IMLS) awarded the University of Richmond's Boatwright Library a two-year, \$478,141 National Leadership Grant to create a digital repository of Civil War-era newspapers.

The Library is digitizing newspapers from both Union and Confederate perspectives from 1857-1865 and will do a cost benefit analysis of the various methods of newspaper content digitization. The project makes digital copies of significant Civil War-era newspapers available online, allows the University of Richmond to build the foundation of an institutional repository, and helps the library and technical worlds identify the best way to provide access to individual articles in digitized newspapers

The IMLS grant funded project brings Tufts (<http://www.perseus.tufts.edu/>) and the Virginia Center for Digital History (<http://www.vcdh.virginia.edu/>) together with the University of Richmond to build a digital repository of Civil War-Era newspapers. Not only will this project provide online access to *The Liberator*, *Philadelphia Ledger* and the *Richmond Dispatch*, but it will also provide a unique research opportunity focusing on the process of digitization and metadata encoding. In addition to the newspapers, it is planned to provide access to a complimentary collections of Richmond related Civil War period resources.

Currently, we are in the process of loading the images and text files of the newspapers into our repository structure. We are using FEDORA as our archive storage and DLXS as our web interface, indexing tool and collection repository. The *Richmond Dispatch* collection spans from November 1st, 1860 to April 3rd 1865. It consists of 4092 page images and 1387 TEI encoded xml files. Later this spring, we will load *The Liberator* and *The Ledger* newspapers. *The Liberator* collection is from January 1861 to May 1865 and consists of 1479 page images, and 368 lightly encoded xml files. *The Ledger* covers from January 1861 to May 1865 and consists of 6096 page images, and 1872 text files from rough OCR results. Access to the newspaper collections is at <http://dlxs.richmond.edu/r/rrdd/>

Completion date for this grant project is September 2006. Our report will not only summarize our research and the outcomes of the Richmond/ Tufts collaboration, but will also compare our results with the other 2 complementary newspaper projects.

The University of Utah, in collaboration with Brigham Young University, will build on an initial collection of three digitized newspapers funded by the Utah State Library with State Library Services and Technology Act funds. This project will expand the collection, distribute the content across four different locations in Utah and aggregate the content into a single, integrated digital collection of historical newspapers.

For more information about the Utah Digital Newspapers go to <http://www.lib.utah.edu/digital/unews/about.html>

The University of Denver, in partnership with the Colorado Digitization Program, Colorado State Library, and Colorado Historical Society will create a 125,000-page newspaper collection from 200 Colorado newspapers that will be available via the Colorado Virtual Library, using Olive software to scan from microform and including online learning modules for K-12 education, and a metadata and indexing system that is based on 19th century natural language.

For more information about Colorado's Historic Newspaper Collection go to <http://www.ColoradoHistoricNewspapers.org>

Completion date for the Civil War Newspaper grant project is September 2006. Our report will not only summarize our research and the outcomes of the Richmond/ Tufts collaboration, but will also compare our results with the other 2 complementary newspaper projects.

Appendix #2: Revised Outcomes Assessment Logic Model

Civil War Newspapers Project—Outcomes Logic Model

Organization Name:	University of Richmond		
Project Name:	A Test bed of Civil War Era Newspapers		
Date Created		Date Reviewed	

Program Influencers <i>(Key entities that help define the program or to whom the program will report results)</i>
<i>Digital library community, U of Richmond Administration, Tufts University and Greg Crane, Historians and teachers, IMLS</i>

Organizational Mission <i>(Organization's mission statement or key action words)</i>

Program Purpose	
We do what? <i>(Summary of key proposed services)</i>	<i>Digitizing Civil War-era newspapers from North and South using cutting edge processes to generate clear, useful images accompanied by consistent, easily searchable metadata and to transfer complementary knowledge between partner institutions</i>
For whom? <i>Target population(s)</i>	<i>The library digitization community so it can adopt new best practices and improve upon those practices. For scholars, students and teachers to have free access to newspapers</i>
For what outcome(s)? <i>(Benefits/changes in skills, knowledge, attitude or life condition.)</i>	<i>Other newspaper projects will adopt and improve our best practices We will establish a repository for 19th century newspapers and Newspapers will be used in university and high school curricula Knowledge (knowledge of what?) will be enhanced between project partner institutions.</i>

Inputs <i>(List items dedicated to or consumed by the program)</i>	Outputs <i>(Program products)</i>
<i>New position Equipment Newspapers Web site Outsource vendors Training consultants Database admin. % of various staff historian tufts staff space</i>	<i># of newspapers digitized Authority file Website DTD's Raw data sets Repository # of images metadata</i>

Program Activities <i>(List key activities needed to provide or manage services.)</i>	Program Services <i>(List services to be delivered directly to participants.)</i>
Digitalization DCR Metadata tagging Authority work Iterative testing Reports – IMLS and more Web design Confer with others Hire for position Purchase computers Establish DTD's	Website best practices Workshop for academics and teachers Access to papers Knowledge exchange

Target Population <i>(List specific characteristics of primary intended participants)</i>
<i>Historians, library digitization community, teachers, students</i>

Intended Outcomes <i>(Changes in skill, knowledge, attitude, behavior, life condition or status)</i>	Indicators (Measures) <i>(Concrete evidence, occurrence, or characteristic that will show the desired change occurred)</i>
Immediate:	
Intermediate:	
Long-term:	

Outcome #1 Digital library technologies peer group will demonstrate knowledge of The Civil War era Newspaper project

Indicator(s)	Data Source <i>(Where data will be found)</i>	To Whom <i>(Segment of population to which this indicator is applied)</i>	Data Intervals <i>(Points at which information is collected)</i>	Target <i>(the number , percent, variation or other measure of change)</i>
The # and % of those who attend conference presentation that articulate 2 project purposes and know one element they can apply to their projects	Presentation evaluation	Conference presentation attendees	Immediate—at conclusion of presentation	50%
The # of sites that link to our repository	WWW	Digital Technologist with repository projects	Every 3 months	5
The # of hits on web site after an announcement of project via a listserv	Web log	Members of listserv	Week after broadcast emails	20

Outcome #2 Digital library Technologists will adopt best practices in future newspaper digitization projects

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
The # of projects that reference any of the project's best practices OR	Survey project managers; Examination of project documentation	Known newspaper digitization projects	May 2005 October 2005 , January 2006 August 2006 , then every 6 months	3
The # and % of staff from other projects who report they were influenced directly by the Civil War Newspaper project	Survey of project managers/staff	– staff involved	May 2005 October 2005 , January 2006 August 2006 , then every 6 months	5

Outcome #3 Historians know about the Civil War Newspaper Repository

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
The # and % of historians who attended the workshops who can name the purpose of the project AND	Workshop evaluation	Those who attend workshop	At end of workshop	100%
The # and % of historians who attended the workshop who revisit the project Web site	Interviews and/or survey	Those who attend workshop	May 2005 October 2005 , January 2006 August 2006 , then every 6 months	80%

Outcome #4 Historians use the Civil War Newspaper Repository

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
The # and % of historians who do at least 1 of the following: <ul style="list-style-type: none"> Incorporate database in a class they teach incorporate in their research 	Interviews and/or survey	Those who attend workshop	May 2005 October 2005, January 2006 August 2006, then every 6 months	50%
The # and % of historians who attended the workshop who report one way in which they have used the repository in their work or research.	Interviews and/or survey	Those who attend workshop	May 2005 October 2005, January 2006 August 2006, then every 6 months	80%

Outcome #5 Project partner Institutions' contributors know new skills and technologies

Indicator(s)	Data Source (Where data will be found)	To Whom (Segment of population to which this indicator is applied)	Data Intervals (Points at which information is collected)	Target (the number , percent, variation or other measure of change)
The # and % of contributors at each partner institution can name 2 new ways the technology can be used or 2 new skills they learned	Interview	Grant participants at all organizations	March 2006 June 2006	100%
The # and % of partner institution contributors use new skills in other projects	Interview	Grant participants at all organizations	March 2006 June 2006	50%
The # or % of contributors that build on skills acquired during project	Interview	Grant participants at all organizations	March 2006 June 2006	25%