

Sample Application

2009 National Leadership Grants for Libraries

Advancing Digital Resources Category

Washington University in St. Louis

The St. Louis Freedom Suits Legal Encoding Project

Abstract

The Washington University Libraries, in partnership with the Missouri History Museum, request \$382,114.95 over a two year period for the remediation of the St. Louis Circuit Court Historic Records Project and digitization of supplementary materials. The primary remediation effort will be the transcription and encoding of these resources in TEI XML. Additionally, we will develop extensions to the TEI for encoding legal documents to reflect legal function, genres, and roles, and employ these extensions in this collection. The resulting extensions and guidelines serve as the basis for a new standard for the encoding of legal documents, and will receive further development as part of the TEI. We believe we have identified a need for such guidelines, and that their establishment and promotion will ultimately encourage more digitization of legal texts, especially historical legal texts, by removing the barrier of local development of complex metadata schema.

With our project partners, within and outside Washington University, we will accomplish the following:

- Digitize, transcribe, and encode the St. Louis Circuit Court Historical Records Project and supplementary materials.
- Develop a set of general metadata and TEI guidelines for the encoding of legal documents.
- Use Named Entity Recognition software to generate personal, organizational, and location names to be ingested into a Resource Description Framework index to map relationships between named entities.
- Develop an interactive component of the project interface that allows users to contribute information, including new names and relationships, to the database.
- Partner with American Culture Studies to accomplish transcription and encoding work on the project, while affording them a valuable educational experience.
- Share institutional areas of expertise with project collaborators.
- Conduct outreach efforts by hosting local, project-related workshops at Washington University and at Missouri History Museum, as well promoting the resources and guidelines at conferences and on listservs.

We look forward to making these important historical materials available for historical research, to further education of slavery and this important period in our nation's history, and to contribute to development of new standards for the larger digital library community. We believe legal historians will also benefit greatly from this project, with the development of the legal encoding extensions.

The cases in the St. Louis Circuit Court Historic Records Project, especially the suits enslaved persons brought against their tacit "owners," are extremely important, and timely. We believe the remediation and expansion of the collection will make these documents significantly more accessible to a wider range of audiences. Creating a full-text searchable collection of these documents and enhancing their use through RDF and user contributed content will provide new levels of meaning for slaves, lawyers, abolitionists, the state of Missouri, and others involved in these cases.

Introduction

The Washington University Libraries request \$38215.95 to remediate and significantly expand, both in content and functionality, the St. Louis Circuit Court Historical Records¹ collection. Part of the increased functionality will be achieved through the encoding of the resulting XML files with extensions to the TEI to reflect the legal function of the documents. The development of these extensions and guidelines for the encoding of legal texts will complement the creation of the digital resource as the other major deliverable of this project.

This collection had its beginnings in the early 1990s, when a historian attached to the Jefferson National Expansion Memorial² in St. Louis, Robert Moore, was alerted to the existence of a number of civil suits as part of the records of the St. Louis Circuit Court in which slaves sued for their freedom. Moore began his research on the cases, which he found covered in coal dust, and wrote the first important article on these freedom suits in 1994³. The first result of Moore's article was the curation of the entire St. Louis Circuit Court records collection by the Missouri State Archives (under the leadership of the then State archivist, Kenneth Winn). The second was two distinct digital projects: the St. Louis Circuit Court Historical Records Project, which consists of 481 cases (a collaboration with American Culture Studies at Washington University) and the Dred Scott project⁴ (a collaboration with the Washington University Libraries, which includes only Scott cases). The former consists of page-images and metadata, while the latter consists of page-images and transcribed HTML documents. These projects have already had a significant use as a classroom resource, as the basis of several scholarly research articles, and most recently, of a Mid-America Emmy Award-winning documentary.⁵

We believe these resources could have an even greater impact through a significant expansion of the collection, the addition of related legal materials and contextual data (including historical city directories) and greater accessibility provided through the transcription and encoding of these resources in TEI XML. A step was taken toward this when, in 2007, the Digital Library Services (DLS) unit of the Washington University Libraries created the Revised Dred Scott project⁶ by updating the interface, converting the HTML documents into TEI XML to make them fully searchable, and extending it to include the full case history. The other element we believed would add significant functionality was a means to represent these documents legal function in XML. Ultimately, an *ad hoc* means of representing this was established for this collection through extension of the TEI (P4). However, our efforts to identify a suitable standard for the encoding of digital legal documents were frustrated, and we believe there is still a need for such a standard.

In addition to the improvements above, legal documents in the collection will also be tagged according to the standard for encoding of digital legal documents, to be developed as a major aspect of the grant. We will take an opposite tack from some previous efforts to develop such a standard, by: 1) developing our standard in close connection with a specific digital project and documents of limited range; 2) adopting an initial modest goal of developing a standard to serve the documents in this collection and similar documents, to avoid the unrealistic aim of creating a standard applicable to *all* legal documents, but 3) creating a framework for long-term extension and development of the standard. The first step in the creation of this framework will be the creation of a TEI Special Interest Group (SIG) dedicated to the development of this standard, with a membership consisting of the committee working on this standard within the project, as well as others interested in its development. We will also use named entity recognition (N.E.R.) software to identify personal, organizational and place-names in the majority of the resources. This data will be exported into RDF/OWL files, in which relationships between all persons in the records can be represented. We will also provide an interface through which users will be able to add new names and relationships to the database, opening the possibility that many of the individuals identified only by first name may be finally identified.

¹ Washington University School of Arts & Sciences. *St. Louis Circuit Court historical records project*. Retrieved January 26, 2009 from <http://www.stlcourtrecords.wustl.edu/index.php>

² National Park Service. (2008). *Jefferson National Expansion Memorial*. Retrieved January 26, 2009 from <http://www.nps.gov/jeff/>

³ Moore, R. Jr. (1993-94). A ray of hope extinguished: St. Louis slave suits for freedom. *Gateway Heritage*. 14, 4-15.

⁴ Washington University Libraries. (2007). *Dred Scott case collection*. Retrieved January 26, 2009 from <http://library.wustl.edu/vlib/dredscott/>

⁵ HEC-TV (Producer), Harris, A., Cudnik, C., Unser, D.C. (Directors). (2008). *Seeking Freedom* [Motion Picture]. United States: HEC-TV cable broadcasting organization.

⁶ Washington University Libraries (2008). *The revised Dred Scott case collection*. Retrieved January 26, 2009 from <http://digital.wustl.edu/d/dre/index.html>

Assessment of Need

Dred Scott Case Collection: A Case Study

Our interest in developing a standard for the encoding of digital legal documents began in 2007, when the DLS unit of the Washington University Libraries created the Revised Dred Scott project.⁷ The Dred Scott resource was initially identified as “low-hanging fruit,” which could be converted into XML quickly. After further evaluation, obstacles to this conversion became apparent. For instance, we realized the “titles” assigned to the documents were primarily descriptive, instead of reflecting the practice in the legal profession of referring to documents by their legal function (*i.e.* “Notice of Motion to Dismiss”).⁸ It became apparent that such issues were due primarily to the lack of a consistent scheme for representing the documents’ legal function, and we began to search for appropriate standards for encoding legal documents in XML.

Assessment

Efforts to bring legal documents into the world of XML generally seem to fall into two categories: efforts by the library and archival community, which have been few and far between, and initiatives by professional groups with a practical interest in legal information. The primary distinction between these efforts seem to be that the former group, exhibits an interest in representing the documents to some extent as artifacts, while the latter group is less interested in the documents *per se*, than in information they contain, or the efficient exchange of that information.

Examples of the latter include LegalXML, the Global Justice XML (GJXML) Data Model,⁹ developed by the U.S. Department of Justice and MetaLex. The first has a focus on electronic submission of court filings. The standard apparently uses XML primarily for its utility in information exchange, as the legal documents themselves are not marked-up, but referenced by the XML as attachments. GJXML similarly has a focus on the exchange of information, not on the encoding of documents *per se*. Of the three, MetaLex demonstrates the greatest concern for representing legal documents (primarily legislative documents) as such, but its aims for universality are achieved by sacrificing granularity. As Winkels, Boer and Hoekstra describe MetaLex: “To achieve *independence of jurisdiction*, the operative principle can only be: *when in doubt, leave it out*. MetaLex is therefore limited to the few features that regulatory documents from these different jurisdictions share.”¹⁰ With the exception of the proposal made by Nick Finke, the majority of digital projects from the academic community (including libraries and archives) treating legal documents either opt for the page-image and metadata collection model or, if using SGML/XML, have only used the markup to reflect a generic text structure, *i.e.*, the encoding does not reflect the legal function of the document. Of those in the former category, where the metadata does reflect legal function of the documents (even to a limited extent) the metadata tends to meet only local interests.

The approach we propose is to avoid these two extremes: of limited, collection-specific metadata for legal documents from the library and archival communities, on the one hand; and of the XML initiatives of the legal community, developed in the abstract which aim at comprehensiveness but either falter in the face of the aim, or achieve comprehensiveness at the expense of specificity, creating a standard unsuitable for libraries or archives. We will avoid these extremes, in effect by combining them. We will develop a standard in close connection with a specific collection, creating a standard that has real descriptive power. But the target for the initial standard will also look beyond the immediate horizon of the local collection, and be designed specifically for extension by encoding projects that adopt it.

Another factor seems to be that encoding communities that have an interest in detailed representation of documents are generally not those who have access to or responsibility for legal documents. (And the St. Louis Circuit Court records are

⁷ Washington University Libraries (2008). *The revised Dred Scott case collection*. Retrieved January 26, 2009 from <http://digital.wustl.edu/d/dre/index.html>

⁸ The St. Louis Circuit Court Historical Records Project does not share this problem, but has limited reference to legal function.

⁹ United States Department of Justice – Office of Justice Programs. *Global justice XML data model*. Retrieved January 27, 2009 from <http://www.it.ojp.gov/jxdm/>

¹⁰ Winkels, R., Boer, A., & Hoekstra, R. (2003). *MetaLex: An XML standard for legal documents*, p. 3. Retrieved January 26, 2009 from http://www.idealliance.org/papers/dx_xml03/papers/02-05-04/02-05-04.pdf

a case in point: they did not become targets of digitization until responsibility for them had been transferred to MSA.) By the same token, those that do serve a community oriented more to a practical, professional interest in documents than to research interests (such as those of legal historians). The key to addressing this circumstance seems to be in effect to introduce these communities, which should both have the effect of more legal documents of scholarly interests being digitized, and of creating a more express need for encoding guidelines. This project is an attempt to take a step in that direction, and our conclusions have been confirmed by our communications with projects dealing digital legal documents.

As we found little guidance for libraries and archives in determining what metadata to capture with regards to legal documents DLS conducted an informal survey of a small number of digital projects and the Scholarly Editing Listserv in December, 2008, regarding the development of a legal encoding standard. The survey consisted of three questions designed to test our main premise that there remains a need for a set of recommendations for the encoding of legal documents. In total, fourteen inquiries were sent out, and we received a 64% response rate (see Supportingdoc4.pdf). The majority of respondents were from Universities, but we also received feedback from two other types of organizations, such as the Church of Latter Day Saints Library and the Center for Tobacco Use Prevention & Research.

- Q1: Only 33% of respondents considered encoding the legal function of the legal documents in the project. We believe this is due to the lack of guidelines. One respondent stated they initially did not consider encoding the legal function of the documents, but felt the need to do so after accruing multiple types of legal materials.
- Q2: Approximately 25% investigated standards for encoding legal documents before beginning their projects, while another 25% did not answer the question. It is likely those who did not answer the question were not part of the original planning. Over half of the respondents, 56%, did not look for an encoding standard. Contributing to this high percentage could be the lack of guidance within particular disciplines.
- Q3: Over half, 56%, answered yes when asked if they believe there is a need for recommendations, whereas 44% did not answer the question. Because many of the projects had not used encoding guidelines or had not investigated standards before beginning their projects, they would not have the experience to recommend that guidelines be established. This could have contributed to the high percentage of those who did not respond.

Overall, responses were very positive, both in survey results and in personal communication with project contributors. Via personal communication, 89% of project contributors expressed interest in our intention to develop a standard for encoding legal documents. While this high percentage of support juxtaposes percentages determined from survey results, this could have resulted from a variety of factors, including some who were not involved in the original project planning and some who, in retrospect, would have used encoding guidelines if they had been available. Those not present for original project planning indicated they would encode project materials using XML, but do not currently have the resources to migrate to a new encoding standard. Of the positive responses for establishing encoding guidelines, three respondents, the Library of Congress, the Digital Docket project and the DocSouth project, expressed interest in establishing guidelines specifically using TEI extensions for legal documents.

The most promising proposal, for our purposes (and those of libraries and archives) for encoding legal documents was an initiative led by Nick Finke¹¹ in 1997 at Text Encoding Initiative's (TEI) 10th Anniversary User Conference. There he proposed a set of legal extensions for the TEI P3 guidelines, but it was not adopted at the time. However, when the Digital Projects Librarian, Cassandra Stokes, presented a poster on the Revised Dred Scott at the Digital Humanities and African American Studies Conference in 2008, one of the sponsoring MITH staff members asked, "When can we expect to see a TEI SIG on encoding legal documents?" This was also the year that the TEI marked its 20th anniversary. The ten-year interim has seen significant changes in the TEI, most especially in the transition from SGML to XML, and more recently, from P4 to P5. The framework has become much more extensible, so new encoding specializations can be more easily accommodated. In preparing this proposal, we have been in informal contact with TEI Council members regarding the possibility of the creation of a TEI SIG on encoding legal documents, and have had a very positive response to this possibility. With improvements in technology, and the maturation of the TEI standard, we believe the time is ripe for a renewed attempt to develop guidelines for encoding digital legal documents with extensions to the TEI.

¹¹ See section Project Resources: Legal Team, page 8.

Conclusion

The fact there have been a number of efforts to develop such a standard in itself bespeaks a need for a standard for the encoding of legal documents. As a result of our experiences, and inquiries, we believe there is a need for such a standard.

National Impact and Intended Results

A standard for encoding legal documents for the library and archival communities:

We will bring a strong team of law librarians, encoding experts and content advisors together to develop a preliminary standard. While its proximate goal will be to meet the needs of our collection of 19th century civil suits from the St. Louis Circuit Court, the guidelines will extend beyond that scope. Not only the result, but also the process of the deliberations will be documented for replication by other digital projects encoding legal texts. Our approach the problem of developing a legal encoding standard will be different from many prior attempts. Instead of aiming for comprehensiveness from the beginning, we will develop a standard of limited scope. But in so doing, we will also provide a model for how the standard can be extended. Well before the completion of the project, the committee developing the recommendation will propose the creation of a TEI SIG for the encoding of legal documents, at which point its development will be opened to the wider community. This SIG will continue as a place for the incubation and further extension of the standard by other projects encoding digital legal documents indefinitely. To widen the applicability of the standard, the TEI XML standard legal encoding standard will be synced to an abstract metadata model for those projects that make their resources available in only a page-image and metadata format. While the immediate impact of the development of the standard will be modest, its potential for information exchange and improved access to legal resources has significant potential.

A timely, interactive and innovative digital resource:

The content of these cases, however, is not simply arbitrary data for the development of an emergent legal encoding standard. On the contrary, these cases—the freedom suits in particular—are extremely important in their own right, and the collection has the potential to offer scholars and the general public a perspective on the institution of slavery and race relations not widely known. For all its notoriety, the Dred Scott case is known in the public imagination more for the resulting U.S. Supreme Court decision than for, *i.e.*, the fact that it was even *possible* for an enslaved person to sue for his or her freedom prior to the Emancipation Proclamation—much less the fact that Dred and Harriet Scott were only two of many who attempted to gain their freedom through the courts. There are, in other words, many heroic stories to be told.

The legal basis for most of these claims was the enslaved person's having lived previously in a free state. As the "gateway to the West,"¹² St. Louis was a crossroads for the Western migration, and many who came to (or through) St. Louis brought slaves with them. Slave owners also often hired out their slaves, including to sites in free states. The stories of these individuals' enslavement and pursuit of freedom told in these records, in fact, unfold across several regions of the country. So, while the jurisdiction of the of the St. Louis Circuit Court was regional, the suits are clearly of national significance and scope, and one important function of this collection will be the disabuse the public of the prevailing belief that slavery was a problem of the South alone and not of the entire nation. This reality will be represented by a GIS component, which will visually depict the roads traveled by litigants to the St. Louis Courtroom.

The original St. Louis Circuit Court Historical Records Project has already had a demonstrated national impact. However, in its current form, the page-images of the hand-written documents are difficult for inexperienced users to read, and the lack of full-text indexing limits the utility to scholars. By removing these barriers to access and expanding the collection to provide additional information about and context for the cases, we believe the impact will be significantly increased. Finally, the creation of RDF/OWL files containing the names of all persons identified in the cases and their inter-relationships, and creating the opportunity for users to interact with the resource by adding new persons and relationships into the database, will both increase use of the resource, and in turn provide information about individuals that scholars have been unable to discover. And these names deserve to be made known. The inclusion of the transcription of the city directories in the project is strategic. While we do not expect to be able to match a significant number of additional litigant

¹² It is fitting that while most know this aspect of the symbolism of the St. Louis Arch, the arch grounds actually incorporate the old courthouse, where the Scott trials were litigated, and in fact, the memorial is also dedicated to Dred and Harriet Scott.

names in the city directories. However, the city directories are one of the most-requested resources at MHM, especially for genealogical research. A significant part of the value of the city directories is to encourage users to use to make use of this resource in adding additional names and relationships to the RDF index in the interactive interface.

A model for digital project development as pedagogy:

Digital projects have long made use of student labor, and have recognized the pedagogical value in such experience. Few, however, take the extra step of formalizing the pedagogical experience in curricular terms. For several years (and largely out of its experience in working with them on the original Circuit Court Records project) American Culture Studies has sent students to work in the Missouri State Archives-St. Louis, where the records are curated, to fulfill the “engaged study” requirement of the program. Students learn curatorial practices and gain experience in archival research, functioning as interns for two thirds of a semester; in the remaining third, they continue their work, and write papers. In the current project, American Culture Studies students will be able to pursue engaged study in connection with their work on the project. Undergraduate students will learn the basics of creating and encoding a significant digital resource, and make important contributions in the correction and higher-level encoding of the legal texts. In the final third of the semester, they will write research papers, and these will also be made available on the project site.

Two graduate research assistants will be given a broad introduction to the digital humanities through the fifth annual summer Digital Humanities Workshop, a program the HDW has conducted in collaboration with DLS. Under the guidance of the primary project advisor, Kenneth Winn, the graduate students will initially contribute to the encoding of case files, and in turn, will train and supervise undergraduate students in this effort for the following two (fall and spring) semesters. While the actual use of student work will not be fundamentally different from prior digital projects, two important differences will be in place: 1) as already stated, this work will have a formal educational component (outside the work of the grant) and in so doing, 2) this will advance the integration of digital humanities into the broader curriculum, for both graduate and undergraduate students. This will follow a model developed locally by Professor Joseph Lowenstein (in his work on a digital archive of the works of Edmund Spenser) who has remarked on the model’s similarity to the lab component in education in the sciences. The parallel to the sciences for this model mirrors the parallel of the promise of the digital humanities, which is to provide humanists (who in generations before, sought to emulate the sciences through method) for the first time with data, which is literally what the process of digitization does. We believe this model for integrating the execution of digital projects with pedagogy has significant potential for emulation in other digital projects, and the role of graduate students as research assistants in particular, will advance the integration of the digital humanities into the wider curriculum as an important, and perhaps essential, tool of 21st century scholarship. Finally, as libraries continue to address the question of their role in the age of Google, one of the viable models is that of libraries as partners, as publishers and co-collaborators in scholarship, with scholars, especially in the humanities. This model of engaged study has the potential to extend that partnership to digital humanities pedagogy as well.

Project Design

Outsourcing of content for imaging and basic encoding:

Final specifications for vendors will be developed prior to the official start of the grant. (This work is not charged to the grant.) In December 2009, transcription and basic TEI encoding of the entire STLCC corpus, plus selected Missouri History Museum (MHM) resources, will be outsourced to vendors. (Two vendors have been contacted who are capable of performing this work, ByteManagers, and Planman Technologies.) The transcription and encoding will be completed by the end of February 2010. In January 2010, DLS will hire and supervise six additional student assistants to digitize and transcribe pages from supplementary source materials from the MSA and image fragile materials from MHM, which will continue through September 2010. The MHM will be shipped to vendors for transcription and basic TEI XML encoding.

Development of standard for encoding legal documents:

December 2009-April 2010, will be devoted to establishing broad guidelines and TEI extensions for encoding legal documents in XML. This will be a collaborative effort between the Washington University Libraries and the Washington University Law Library. Dr. David Konig, Washington University History Professor, and two others, John Barden and Nick Finke, will act as legal advisors to the group. During this time, DLS will provide project partners with XML and TEI training. In November 2010, a team member will attend the 2010 TEI Members Meeting to request the creation of a SIG

for the development of TEI extensions for encoding legal documents; once approved, further development of the TEI extensions and guidelines will come out of the SIG. In addition to the extensions, this group will also develop a tutorial.

Higher level encoding of expanded of collection documents:

In March 2010, DLS student assistants will begin to correct the transcribed TEI files. NER software (GATE) then will be run on the TEI XML files to identify and encode personal, organizational and place-names. A period of training the software will be required to improve accuracy. Following achievement of acceptable results (60-70%) student workers will correct the processed files from April-May 2010, at which point the initial specification for the encoding of legal texts will be ready. From May 2010-September 2011, students and student workers will perform higher-level TEI XML encoding, including the TEI extensions for legal document encoding. From May-August, 2010, a model for the RDF/OWL files will be developed, relying on existing ontologies such as W3C Time, CIDOC-CRM and SKOS. We are using the Henry III Fine Rolls¹³ project as a model for this work. In early 2011, DLS will extract named entities from the TEI XML corpus into RDF/OWL files, and encode relationships (RDF triples) between the named entities. Washington University Libraries' Metadata Librarian, Tim Lepczyk, DLS Metadata Librarian, will supervise this phase of the project.

Development of pedagogical component:

In April 2010, Ken Winn will lead development of the "engaged study" component. The fifth annual Digital Humanities Summer Workshop in June 2010 will inaugurate the program for the 2010-2011 academic year. The workshop will offer digital humanities training for two Graduate Research Assistants who will assist Winn in managing the program, and the participating undergraduate students working on the project.

Development of delivery and interactive components:

In December 2009, the HDW's Steve Pentecost and library staff will collaborate on the design and construction of the system and user interface for the new collection. This will include customization of the search interface for DLXS, the implementation of the FEDORA system to house the RDF index, and the interactive user component (for adding named entities and RDF relationships). Development will continue through July 2011. Responsibility for the GIS component of the site will fall under the Libraries' GIS Librarian, Cheryl Morton. Design of the web site will be led by DLS staff-member Shannon Showers. A beta site will be ready by July 2011, with the interactive elements in place for usability testing to take place through September 2011, led by the Libraries' Assessment Coordinator, Carol Mollman. September-October 2011 will be devoted to the analysis of results and making changes ahead of the site debut November 2011.

Development of outreach plans and dissemination:

Building on planning done in the prior year, in June 2011, planning will begin on a set of programs for outreach that include co-hosting lectures and workshops with the St. Louis-area Public Libraries, the Missouri History Museum, and the National Park Service (Jefferson National Expansion Memorial). Three subject librarians, Barb Rehkop (Government Documents) and Rudolf Clay (African-American Studies) will design and host the programs with project team members.

Evaluation Plan

Washington University Libraries' Assessment Team, led by the Libraries' Assessment Coordinator, and comprised of several, specialized subgroups, will use evaluative processes throughout the grant period and after project completion. Evaluation will help measure the project's success in meeting the following project goals:

1. Increase visibility and use of the St. Louis Circuit Court Records

An enhanced project interface, increased functionality and added content, including more than 1,700 documents held by the Missouri State Archives (MSA) and over 1100 by Missouri History Museum (MHM) will increase use of the Circuit Court Records Project by new and existing researchers. WebTrends, a log analysis tool, will be used to compare usage of the current with that of the revised site. Tracking citations of the project in scholarly publications and the number of web sites linking to the project web site will also indicate increased usage. Citations will be tracked using citation and journal

¹³ The Henry III Fine Rolls Project similarly has a TEI XML encoded collection and an associated index of RDF/OWL files. Retrieved January 28, 2009 from <http://www.frh3.org.uk/cocoon/frh3/content/about/technical.html>.

databases and search engines and will be a long-term evaluative project. The number of external web sites linking to the project interface will provide short-term evaluation and will be captured with Google Webmaster Tools.

2. *Encourage collaboration and integration of user contributed content*

Added functionality, including full-text search and the ability to contribute content, is expected to lead approximately 5% of visitors to contribute data via the wiki interface, which may include identification of people and locations or genealogical information. User-supplied data will be identified separately from metadata created by DLS and will be compared with the number of web site visitors captured via log analysis to determine a percentage. Usability tests will be conducted on the project's beta web site by the Assessment Team's Usability subgroup and suggestions for improvements to the interface, including the wiki and RDF components, will be evaluated before the public launch.

3. *Broaden and deepen access to source material using Resource Description Framework (RDF)*

Through RDF, researchers will be able to draw new conclusions from project materials, identify and evaluate relationships, and track migration patterns of slaves. Citations and scholarly articles on project related topics will be evaluated for effects RDF generated relationships and subsequent conclusions have had on research. In addition, known researchers will be surveyed a year after project completion to evaluate how full text searching and RDF have enhanced research of the Circuit Court Records documents. Researchers will be identified by a registration component on the project interface, which will include a user's name, role (researcher, student, etc.), and interest in the project material. Registration will help gather demographic data to ensure that DLS is meeting the needs of its identified target audiences.

4. *Increase availability of primary documents for use by educators*

This project will aid Missouri teachers and students in meeting state grade level expectations for elementary school, middle school and high school, such as understanding the causes and consequences of the Civil War and Reconstruction and Missouri's role as a border state. The number of registered teachers and students visiting the project interface will be a measure of success in helping educators and students meet grade level expectations outlined by the Missouri Department of Elementary and Secondary Education. Partnership with the Jefferson National Expansion Memorial, which includes the Gateway Arch and the Old Courthouse, and MHM, will allow DLS to survey students and teachers who visit these important St. Louis landmarks on field trips for the purposes of measuring the project's influence on their education.

5. *Share expertise among project collaborators and gain knowledge through partnerships*

Project collaborators, including MHM, the Washington University Law School Library, and students working on the project, will benefit from sharing institutional areas of expertise, including guidance in legal terminology and a series of DLS hosted workshops on XML and project related markup. DLS has hosted several XML workshops in the past and will employ previously used surveys before and after workshops to evaluate participants' success in learning the material. Washington University librarians in turn will learn about engagement with various public user communities from MHM.

6. *Develop and share legal encoding guidelines for use in other digital projects*

Following the transfer of the initial digital legal document encoding standard to a TEI SIG, the legal encoding guidelines will be freely available for use, comment, and extension by the TEI community and will act as a model for libraries and archives in digitizing legal documents. The guidelines will be considered initially successful when accepted by the TEI and when a substantial number of digitization projects use the guidelines. Digitization projects listed in the project bibliography were surveyed to gauge the need for legal encoding guidelines and will be surveyed again, after the guidelines are complete, to assess the influence these standards have had on those projects' encoding practice.

Evaluation will begin at the outset of the project and continue throughout the duration of the grant period. Following the launch of the site, user contributed content and registration will be evaluated. Evaluation will continue after project completion with continuing outreach efforts to classrooms, libraries, and visitors to the Jefferson National Expansion Memorial and the Missouri History Museum. In addition, the process of surveying researchers and determining effects on scholarly research will be an ongoing evaluative project, allowing the library to gauge long-term impact.

Project Resources

Budget

The Washington University Libraries request a grant of \$382,15.95 to develop extensions to the TEI for encoding digital legal documents and simultaneously apply that encoding to an expanded and remediated iteration of the St. Louis Circuit Court Historical Records Project. We believe that the creation of these extensions (and the accompanying general metadata model) not only will meet an immediate need, but will spur libraries and archives holding important legal material to digitize materials with the help of these standards. The total cost of the project is \$767,445.95 however, that figure does not reflect the majority of server maintenance and other IT support costs for the project. We will meet cost-share requirements through contributions of time from eighteen staff members from across the University and MHM.

Project Director—Andrew Rouner. Dr. Rouner is the founding Director of the Digital Library at Washington University. Since coming to Washington University, he has guided and assisted on several projects, including *Eyes on the Prize* transcripts, *Unreal City* and the forthcoming *Red Brush: Women Writers of Imperial China*, a collection of over 700 brief Chinese texts in XML. Prior to Washington University, he was Digital Resources Librarian for the IMLS-funded *Richmond Daily Dispatch* project, which was a collaboration between the University of Richmond and the Perseus Project. He will spend 20% on this project, and will, along with Cassandra Stokes, serve as the primary coordinator.

Project Advisor—Kenneth Winn. As a historian and the former State Archivist of Missouri, under whose leadership the original St. Louis Circuit Court Historical Records Project was created, Dr. Winn brings intimate knowledge of the material and its context. He will serve as advisor to the project in all aspects, and will also serve in both pedagogical and managerial roles in guiding graduate research assistants and undergraduate students in contributing to the encoding and research work on the project. He will also have a part-time faculty appointment in American Culture Studies.

Legal Team

A strong committee has been organized to develop a standard for encoding legal documents, which brings together members with legal and encoding expertise in equal numbers. Members from the encoding side include Andrew Rouner, who will lead the committee, Cassandra Stokes, Tim Lepczyk and Perry Trolard, who recently worked on the Mellon-funded development of the TEI "Tite" schema under the leadership of John Unsworth. Legal expertise on the regular committee includes four Washington University Law Librarians: Dora Bertram, Hyla Bondareff, Wei Luo and Aris Woodham. During December 2009-April 2010, the committee will meet every two weeks, with bi-monthly meetings scheduled thereafter to coincide with the project's monthly stakeholder meetings. The committee will rely on electronic communication and management of files created, and will use a content management system for this purpose.

Three outside advisors will consult with the committee throughout the process: Nick Finke, John Barden, and David Konig. All three bring not only extensive legal experience, but significant engagement with the history of slavery. Nick Finke's positions have included tenure as Director of the Center for Electronic Texts and the Law at the University of Cincinnati College of Law Library in Ohio and as librarian and counsel for the National Underground Railroad Freedom. John Barden is currently Director of the State Law and Legislative Reference Library for the Maine State Legislature. In addition to his law and information science degrees, he also has a doctorate in history, and wrote his dissertation on the John Carter Nomini Hall emancipation. David Konig is a renowned legal historian currently teaches at Washington University as a Professor of History and a Professor of Law. Konig was involved with the original Circuit Court Historical Records Project and has written papers based on the archive.

Digital Production Team

The project will rely on the technical and project management skills of the DLS staff Cassandra Stokes, Tim Lepczyk, and the HDW's Perry Trolard, and the MHM's Emily Jaycox. This team will supervise various efforts throughout the project, including digitization and transcription, development of delivery system, higher-level encoding of collection documents, creation of RDF files, and dissemination. Monthly stakeholder meetings scheduled throughout the project will be devoted to planning the tasks. The committee will also rely on electronic communication and project management, for which the Libraries has a dedicated content management system (Microsoft's SharePoint) in place. Assisting in these efforts will be DLS' Digital Projects Assistant Shannon Showers, American History Subject Librarian Makiba Foster, GIS Librarian Cheryl Morton, HDW's Digital Humanities Specialist Steve Pentecost, and MHM's Assistant Librarian Jason Stratman. Having collaborated successfully in the past with the HDW and MHM, DLS is confident that continued partnership with these two organizations will provide opportunities for sharing institutional expertise. However, we will not rely on

serendipity, but have built in a number of specific opportunities in the project plan for all project partners to learn from each other, including collaborations on the legal committee, outreach activities, and formal training sessions.

Outreach Team

Outreach will be a significant component to the project, and DLS has a team with significant outreach and instruction experience, both in public and academic libraries, to plan and implement outreach efforts with St. Louis-area Public Libraries, the National Park Service and the Missouri History Museum. Emily Jaycox will lead the committee, along with Rudolph Clay, African and African-American Studies Subject Librarian, and Barbara Rehkop, Government Information Librarian. The committee will work independently and report progress at the monthly stakeholder meetings.

Dissemination

Our dissemination strategy will largely mirror our strategy for the development of the legal encoding guidelines: while our ultimate aim will be national (and international) exposure, we will begin locally, and build out. In addition to the usual means for promoting a digital collection, we plan to host a number of workshops and informal talks locally. On the one hand, we will use these as opportunities to solicit feedback on the resources in development, the results of which will be fed back into project development. On the other hand, we expect to learn from workshop attendees of further outreach and dissemination opportunities, *i.e.*, which online resources they use most, and what other organizations would be interested in learning about the project. These in turn will be targeted for announcements and demonstrations. These workshops will be lead by two subject librarians from Washington University: Barbara Rehkop, Government Documents Librarian (and formerly of the St. Louis Public Library) and Rudolph Clay, whose specialization is in faculty outreach and African-American Studies. Beginning in May of 2011 we will begin coordination on two workshops and a series of informal talks with MHM, as well as with the St. Louis Public Library and the National Park Service. The workshops taught with MHM will target people interested in family history, and students and scholars involved with the digital humanities, respectively.

The first workshop will demonstrate to users how to create representations of family members and associate them with persons named in the RDF index. Also, we will instruct users on how to create their own online content that represents their family history using web-services like Google and Flickr. The orientation of the workshop fits well with the MHM mission to welcome the public community into the creation of digital content. The second workshop will focus on students, scholars and educators from the region and show how the collection can be used in the classroom. We will also lead a series of informal talks on the collection to area residents at different branches of the St. Louis Public Library. Working with the Research Collections Manager, Joe Winkler, we will develop talks that are informative and have broad appeal to the user communities of the St. Louis Public Library. Each workshop and talk will be evaluated and used to inform future activities. Finally, we will develop a workshop with the National Park Service in St. Louis the series of events to commemorate the American Civil War sesquicentennial.

We will also promote the resource with postings to listservs in the library and encoding community, such as the TEI listserv, D-LIB as well as in law librarian, and academic listservs such as H-NET. We will promote both the resource and the legal encoding guidelines with presentations and posters at conferences, including the TEI Members' Meeting and the Chicago Digital Humanities Conference. We will also approach OCLC on producing a webinar on the use of the guidelines (following the grant period).

We will promote the resource through the exposure of Dublin Core metadata through harvesters, including that of the IMLS, OAIster and the State's *Missouri Digital Heritage* web site. We are currently in the process of addressing the inability of major search engines to index the "deep web"¹⁴ and will employ the same strategy with the remediated Freedom Suits site. Using the Revised Dred Scott collection as a test, we have established a workflow to automatically create static web-pages for each text as well as an XML sitemap, which we have submitted to Google for indexing.¹⁵

¹⁴ Hagedorn, Kat. Summer 2008. Google Still Not Indexing Hidden Web URLs. *D-Lib Magazine*. Retrieved 01/20/09 from <http://www.dlib.org/dlib/july08/hagedorn/07hagedorn.html>.

¹⁵ Deridder, Jody. 3/24/2008. Googilizing a Digital Library. *The Code4Lib Journal*. Retrieved 01/20/09 from <http://journal.code4lib.org/articles/43>.

Sustainability

In order to increase adoption of the legal encoding guidelines, we will apply to the TEI for the formation of a SIG legal encoding at the TEI Members' Meeting in November 2010. The SIG will initially consist of project members but will also be open members from the TEI. By transferring the legal encoding guidelines at an early stage to the TEI, we will assure conformity of our encoding to the public standard, and use the grant period to promote the guidelines. Because many projects will continue to choose to mount collections as page-image and metadata collections, we will also sync the TEI guidelines with an abstract metadata model (expressed in METS and Dublin Core). We believe, as was suggested to us by a TEI member, that what finally may be needed would be "more than a SIG, in actual fact an entire community of practice encoding legal documents with a TEI extension...[such as] the EPIDOC¹⁶ initiative," which is our ultimate goal.

DLS is a relatively new unit of the University Libraries. While there was a primary concern to begin development of collections, we balanced this need against the need to address long-term infrastructure and policy issues. To accomplish this, we converted the library-wide committee previously charged with developing local digital content, the Digital Library Team (DLT) into a steering committee with oversight of three subcommittees, charged respectively with 1) the establishment and promotion of local metadata standards (the Metadata Committee);¹⁷ 2) the development of policies for the selection and support of locally developed digital collections, including rights and long-term maintenance of digital assets (the Digital Projects Selection Committee; and 3) the creation of a digital asset preservation system (the "Dark Archive" Committee). Each committee supports the other to promote the use of standards at Washington University, which in turn facilitates the effective curation of digital assets and projects.

We recognize that both preservation of digital assets, and continued access to digital resources are significant concerns of content creators, but also that such promises cannot be made lightly. We have therefore developed a SIP (submission information package) agreement that outlines rights and responsibilities to content creators. Because technology changes, we have chosen migration as a preservation strategy to ensure long-term access to content. We are currently in the process of developing a FEDORA-based system to catalog and preserve essential digital assets. The first candidate for this system will be the St. Louis Circuit Court Historical Records project, which we have committed to curating regardless of the outcome of this project. Should our application be successful, however, we see the collection as a platform for almost limitless extension. One of the benefits of digital collections is that it can combine assets from diverse institutions, as we are doing with resources from MHM and MSA. We intend the partnership(s) from the project itself to continue long after the completion of the grant, and envision the continued addition of significant resources, as well as new services and means of accessing the content well into the future.

Conclusion

The St. Louis Freedom Suits Legal Encoding Project will significantly expand the collection while developing new metadata models, TEI extensions and guidelines for encoding legal documents. Driven by the experience digitizing the Dred Scott Case Collection, DLS has identified a need for these guidelines both for Washington University Libraries and the larger digital library community. Although there have been previous attempts to develop legal encoding guidelines, Washington University Libraries believes that with a specific set of documents and a narrow focus, the guidelines developed as part of this project will become an accepted standard for other legal digitization projects. The addition of new technologies such as NER and RDF will provide additional functionality and research opportunities for exploring these important resources. Partnership with colleagues at the Missouri History Museum, the Washington University Law School, the Humanities Digital Workshop and the American Culture Studies department will ensure best practices for digitization methods, developing legal guidelines, and will encourage the development of digital humanities scholarship. We anticipate this project will have a wide-ranging and lasting impact for research and digital library development.

¹⁶ Retrieved 01/26/2009, from <http://epidoc.sourceforge.net/>, an expanded TEI initiative to encode classic inscriptions and papyri.

¹⁷ Retrieved 01/22/2009, from http://digital.wustl.edu/standards/meta_required.html

BUDGET FORM - PAGE FOUR

Section B: Summary Budget

	\$ IMLS	\$ Cost Share	\$ TOTAL COSTS
1. Salaries and Wages	122,421.00	230,692.00	353,113.00
2. Fringe Benefits	7,133.54	55,527.00	62,660.54
3. Consultant Fees	1,000.00		1,000.00
4. Travel	14,000.00		14,000.00
5. Supplies and Materials	50,771.00		50,771.00
6. Services	96,671.00		96,671.00
7. Student Support			0.00
8. Other Costs			0.00
TOTAL DIRECT COSTS (1–8)	291,996.54	286,219.00	578,215.54
9. Indirect Costs	84,429.06	100,176.65	184,605.71
TOTAL COSTS (Direct and Indirect)	376,425.60	386,395.65	762,821.25

Project Funding for the Entire Grant Period

1. Grant Funds Requested from IMLS	376,426.00
2. Cost Sharing:	
a. Cash Contribution	
b. In-Kind Contribution	386,395.65
c. Other Federal Agencies*	
d. TOTAL COST SHARING	386,395.65
3. TOTAL PROJECT FUNDING (1+2d)	762,821.65
% of Total Costs Requested from IMLS	49.45%

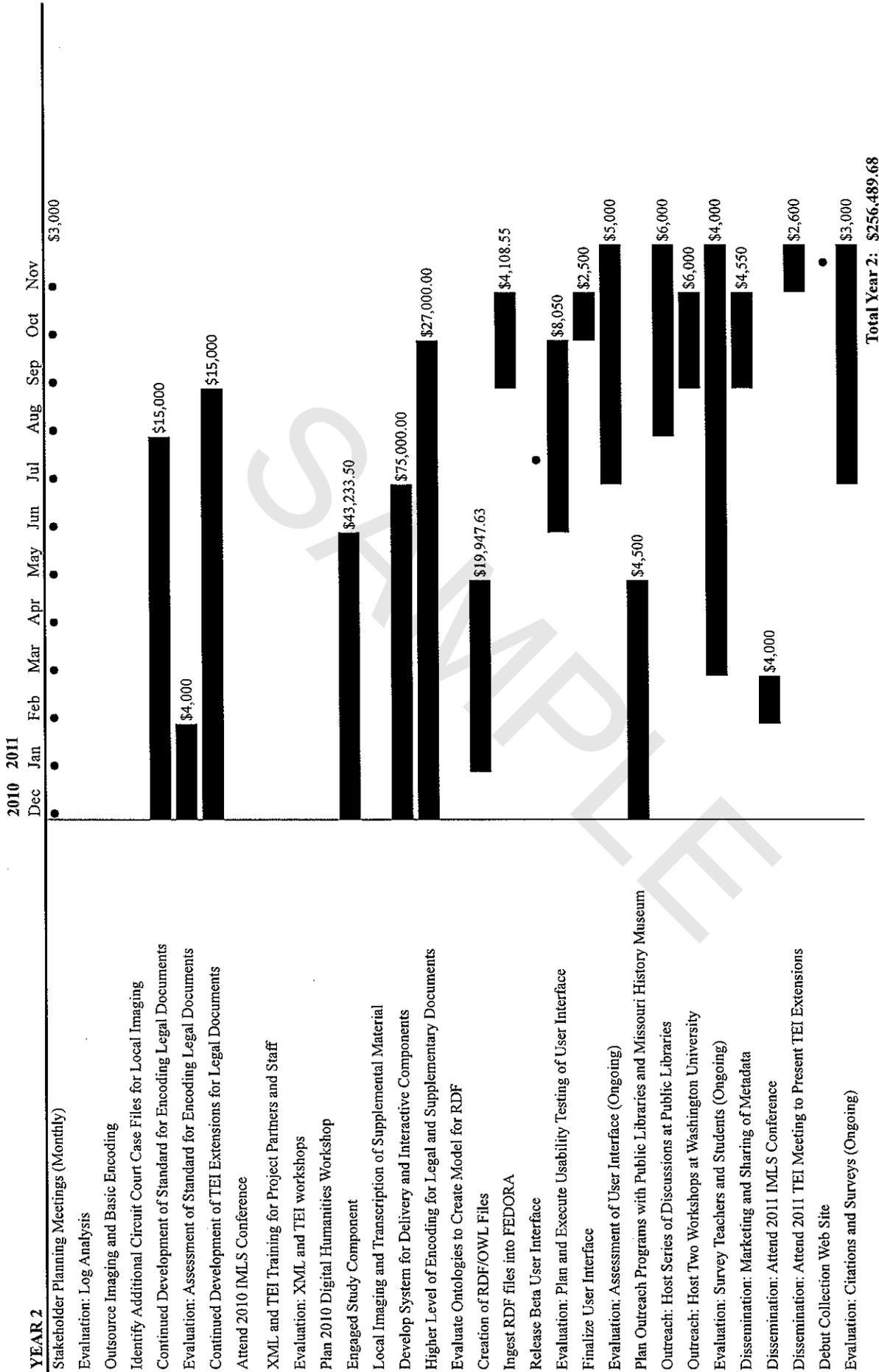
* If funding has been requested from another federal agency, indicate the agency's name:

Schedule of Completion: St. Louis Freedom Suits / Legal Encoding Project

YEAR 1	2009 2010											
	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Stakeholder Planning Meetings (Monthly)												
Evaluation: Log Analysis	\$2,500											
Outsource Imaging and Basic Encoding			\$96,670.70									
Identify Additional Circuit Court Case Files for Local Imaging			\$3,000									
Develop Standard for Encoding Legal Documents					\$50,000							
Evaluation: Assessment of Standard for Encoding Legal Documents												\$8,000
Develop TEI Extensions for Legal Documents					\$15,000							
Attend 2010 IMLS Conference				\$4,000								
XML and TEI Training for Project Partners and Staff				\$6,000								
Evaluation: XML and TEI workshops				\$1,595								
Plan 2010 Digital Humanities Workshop						\$2,200						
Engaged Study Component												\$43,233.50
Local Imaging and Transcription of Supplemental Material										\$130,509		
Develop System for Delivery and Interactive Components												\$75,042.22
Higher Level of Encoding for Legal and Supplementary Documents												\$56,845.85
Evaluate Ontologies to Create Model for RDF												\$12,360
Creation of RDF/OWL Files												
Ingest RDF files into FEDORA												
Release Beta User Interface												
Evaluation: Plan and Execute Usability Testing of User Interface												
Finalize User Interface												
Evaluation: Assessment of User Interface (Ongoing)												\$1,000
Plan Outreach Programs with Public Libraries and Missouri History Museum												
Outreach: Host Series of Discussions at Public Libraries												
Outreach: Host Two Workshops at Washington University												
Evaluation: Survey Teachers and Students (Ongoing)												
Dissemination: Marketing and Sharing of Metadata												
Dissemination: Attend 2011 IMLS Conference												
Dissemination: Attend 2011 TEI Meeting to Present TEI Extensions												
Debut Collection Web Site												
Evaluation: Citations and Surveys (Ongoing)												

Total Year 1: \$510,956.27

Schedule of Completion: St. Louis Freedom Suits / Legal Encoding Project



Specifications for Projects that Develop Digital Products

Part I. Complete the appropriate sections:

A. Converting Non-Digital Material to Digital Format

A1. Describe types and original formats of materials to be selected for digitization and quantity of each. We will image select pages from 34 record books (1750 images) 6 indices (100 images) and 24 city directories (8134 images); of these, only the city directories will be outsourced to a vendor for imaging. Vendor will also create 64 basic TEI XML files from these images.

A2. Identify copyright issues and other potential restrictions with regard to the original non-digital material.

Public domain: 100 % of total Privacy concerns: % of total.
Plan to address:

Permissions have been obtained: 85 % of total Other: % of total. Explain:

Permissions to be requested: % of total

Plan to address: Exact references from case files to page numbers on record books will be determined following transcription of cases; we will not know in advance if some volumes will need of preservation prior to digitization.

A3. Describe how the newly digitized material will be made available to the public. Explain the terms of access and conditions of use. Identify and explain any restrictions that will apply to digitized material, and specify what percentage if any of the total material will be subject to restrictions. The content will be available to the public through the Internet. Users are freely able to access and use the content. Government documents in the collection have no copyright or use restrictions. Publication of other material may only be used for educational purposes (20%).

A4. List the equipment and software, with specifications, whether purchased, leased or outsourced, that will be used (e.g., camera, scanner, server, A/D audio or video converter):

Locally imaged material will be shot with Canon Digital Rebel XSi, Bogen - Manfrotto Repro Copy Stand with Large Baseboard 30x36, Bogen Imaging Lighting Assembly with Four Reflector Heads, computer files created in Adobe Photoshop CS4.

B. Repurposing Existing Digital Content

B1. Describe types and original formats of digital materials to be selected for repurposing and quantity of each.

TIFF images: 16316 (1924 from city directories, 11062 from case files, 3330 from freedom bonds and licenses); XML files: 114 (111 case files from Revised Dred Scott; 1 city directory XML file, 1 corrected text file of The Colored Aristocracy of St. Louis).

B2. Identify copyright issues and other potential restrictions with regard to the original digital material.

Public domain 100 % of total Privacy concerns: % of total.
Plan to address:

Permissions have been obtained: 100 % of total

Permissions to be requested: % of total Other: % of total. Explain:
Plan to address:

B3. Describe how the repurposed material will be made available to the public. Explain the terms of access and conditions of use. Identify and explain any restrictions that will apply to repurposed material, and specify what percentage if any of the total material will be subject to restrictions. The content will be available to the public through the Internet. Users are freely able to access and use the content. Publication of the material may only be used for educational purposes. Government documents in the collection have no copyright or use restrictions.

B4. List the equipment and software, with specifications, whether purchased, leased or outsourced, that will be used (e.g. MPEG encoder, non-linear editing system, GIS software).

XML files will be created using Oxygen and EmEditor on Dell workstations.

C. Creating New Digital Content

C1. Describe types of materials to be created in digital form and quantity of each.

The hand-written content from the 11062 TIFFs of case files (above) will be transcribed and encoded in (new) basic TEI XML by a vendor. An additional 1117 XML files will be created from pre-existing images of freedom bonds and licenses. We will create

50 RDF/OWL files of named entities extracted from person, place and organizational names identified in the case files. 1 GIS map layer tracing the routes slaves took entering and leaving St. Louis will also be created.

C2. Describe plan to obtain releases/permissions from project content creators and subjects. No permissions will be required; the only logistical barrier is that a fraction of material may require conservation before imaging.

C3. Describe disposition of ownership and use rights of new product. Describe how the new product will be made available to the public. Explain the terms of access and conditions of use. Identify and explain any restrictions that will apply to new content and specify what percentage if any of the total material will be subject to restrictions. The content will be available to the public through the Internet. Users are freely able to access and use the content. Publication of the material may only be used for educational purposes. For users to add content to the site they will need to complete a short registration process.

C4. List the equipment and software, with specifications, whether purchased, leased or outsourced, that will be used (e.g., camera, audio recording equipment, video recording equipment, encoding software, server). In order to create the RDF/OWL files we will use GATE for the Named Entity Recognition software, and Protégé to edit the files and create supplementary ontologies. The RDF/OWL files will then be stored using the Mulgrara triplestore index in a FEDORA repository. To create the GIS map, we will use Arc GIS.

Part II. Answer all questions:

5. Specify each type of file format (e.g., TIFF, JPEG, MPEG) to be produced and anticipated quality (e.g. minimum resolution, depth, tone, pixel dimensions, file size, sampling rate) of each.

Master: 27,300 uncompressed tiffs at 600 dpi, 24-bit color, and a file size approximately 47 MB.

Access: 27,300 jpeg2000 at 300 dpi, 24-bit color.

Thumbnail: We will not use thumbnails for this collection.

6. Describe the delivery medium that will be used (e.g. Internet, broadcast, DVD). Content will be served via the Internet.

7. Describe the underlying software to manage and/or present the content (e.g. DSpace, Fedora, ContentDM). We will use DLXS to index and manage the TEI XML files, Fedora to preserve all of the files and to index and manage the RDF/OWL files, and ESRI GIS Portal to serve the GIS map.

8. Describe the quality control plan. There are four areas where we will perform quality control. The first area is content created by the vendor. The transcriptions performed by the vendor will use a double blind method keyboarding. Once we receive these files, student workers will correct the files as needed. Supervisors will make random searches across files to assure consistency periodically. Newly imaged TIFFs will be checked for corruption through transformation of the images into access JPEG2000. If corrupted TIFFs are found, they will be re-imaged locally. The next area for quality control is documents that are locally imaged and transcribed. Before each imaging session, we will check to make sure the calibrations and settings are correct on the equipment. Images will then be cropped by student workers under the supervision of the Digital Projects Assistant. Transcriptions and creation of TEI files will be performed using a template created by the Metadata Librarian.. For the higher level encoding, the Metadata Librarian will validate files against the project's document type declaration (DTD). The third area of quality control is to configure and test the Named Entity Recognition software. The Humanities Digital Workshop programmer working with DLS staff will perform a series of tests on sample text until we get the acceptable results. The last area of quality control will be on the usability of the site and of the interactive component. We will perform usability testing and assesment, correcting any problems which are raised.

9. Explain how descriptive and administrative metadata will be produced and used to describe and manage the content. Include the standards that will be used for data structure, content (e.g. thesauri), protocols, preservation and administrative information and communication of the content (e.g., MARC, EAD, Dublin Core, PBCore, VRA Core Categories, or Categories for the Description of Works of Art).

Descriptive metadata for the city directories will be transformed using XSLT to crosswalk MARC into TEI. Basic metadata for the court records will be derived from the project site, and additional metadata added in the TEI Header. Additional descriptive metadata will be gathered using Named Entity Recognition software to encode people, places and organizations. Also, the court cases will be encoded in TEI using the legal guidelines, which we will develop as part of the project. We will use Library of Congress authority headings for subject terms in the metadata. The TEI files will be converted into OAI records to share with

other institutions and also into qualified Dublin Core for the Missouri State Library's, Digital Heritage site which uses ContentDM. Entities encoded from the NER software will be exported into RDF/OWL files and indexed in a triplestore using Fedora. Archival versions of the TEI files will be saved with their corresponding images using METS.

10. Describe plans for preservation and maintenance of the digital files during and after the expiration of the grant period (i.e., storage systems, migration plans and commitment of institutional funding).

The preservation and maintenance of the digital files for this project will be follow the preservation guidelines and actions that Digital Library Services has established for all digital production. In the production phase, content will be stored online through a SAN, which is regularly backed up and managed by our Systems Operations and Support department. When production is complete, we will have archival TIFFs and XML files saved in a Fedora-based repository for digital content.

Washington University Libraries recognizes that both preservation of digital assets, and continued access to digital resources are significant concerns of content creators, but also understands that such promises cannot be made lightly. We have therefore taken a two-track approach to the issue, and have developed a SIP (submission information package) agreement modeled after those used by the California Digital Library and the University of Michigan, that outlines rights and responsibilities of content creators and curators alike. Because technology changes, we cannot promise resources will be delivered in exactly the same way in the long term. But where projects have followed standards for content and metadata creation, the digital assets can be preserved in perpetuity. We are currently in the process of developing a FEDORA-based system to catalog and preserve essential digital assets such as archival TIFF images and valid XML (adopting a policy of forward migration of assets). While certification on the RLG/NARA trusted digital repository checklist is not in our immediate plans, it is a long-term goal.

11. If content will be provided on the Internet, indicate agreement to submit collection level records for digital products to the IMLS Digital Collections Registry. State reasons for selecting alternative approaches.

We will submit collection level records to the IMLS Digital Collections Registry.

12. Provide URL(s) for applicant's previous digital products, if applicable.

Please see the attached document.

Attachment:

12. Provide URL(s) for applicant's previous digital products, if applicable.



Eyes on the Prize Interviews

[<http://digital.wustl.edu/e/eop/index.html>]

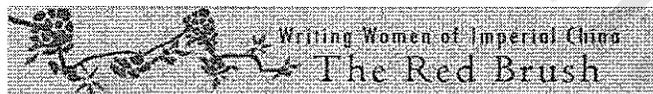
Eyes on the Prize provides full-text access to transcripts of interviews filmed for the PBS Eyes on the Prize documentary series, produced by Blackside, Inc. Founded by Henry Hampton, it was the largest African-American owned film production company of its time. The interviews, which provide accounts of the American Civil Rights Movement, are taken from the first 6 episodes of the series. The larger Henry Hampton Collection resides in Washington University's Film and Media Archive, which DLS partnered with in developing this collection.



Unreal City

[<http://digital.wustl.edu/w/whi/index.html>]

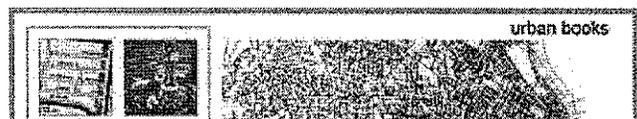
In partnership with the Missouri History Museum, Washington University Libraries digitized approximately 3,300 St. Louis fire insurance maps produced by Alphonso Whipple between 1870 and 1898. These maps provide the most complete visual record of the evolving built environment in St. Louis. Previously restricted in use by their brittle condition, the maps are now available to the public through this joint digital project. In addition to full-text search and images, the project interface includes an interactive, GIS component.



The Red Brush

[<http://digital.wustl.edu/r/red/index.html>]

The Red Brush is a virtual companion to the anthology, *Red Brush: Women Writers of Imperial China* co-authored by Wilt L. Idema, Professor of Chinese Literature at Harvard University, and Beata Grant, Professor of Chinese Language and Literature at Washington University in St. Louis. The collection contains over 500 Chinese language texts, by and about women, that are translated and discussed in *Red Brush*. The interface provides full-text search in both English and Chinese. This project represents a significant partnership between Washington University Libraries and Washington University faculty.



Urban Books

[<http://digital.wustl.edu/u/urb/index.html>]

Digital Library Services worked on this project in collaboration with Jana Harper, an instructor at Washington University's Sam Fox School of Art, and the Special Collections unit of University Libraries. The collection contains books used as original study material for the School of Art's Urban Books class. Using previously generated metadata and digital images, DLS created fully searchable TEI files for enhanced usability. The collection is comprised of professional artists' books, as well as students' final projects, and now numbers over 150 books.

PARTNERSHIP STATEMENT

Complete one of these forms for each formal partner.

Legal name of applicant organization (5a from Face Sheet): Washington University

1. Legal name of partner organization: Missouri History Museum

2. Partner DUNS number: 156197303

3. Mailing address:

Street1: P.O. Box 11940 Street2:

City: Saint Louis State: MO Zip+4: 63112-0040

4. Partner Web address: <http://www.mohistory.org>

5. Partner project contact name: Emily Jaycox

Title: Librarian

Telephone number: 314-746-4508 E-mail: emily@mohistory.org

6. Governing control of partner (choose one):

- | | |
|---|--|
| <input type="checkbox"/> State Government | <input checked="" type="checkbox"/> Nonprofit with 501(c)3 IRS Status (Other than Institution of Higher Education) |
| <input type="checkbox"/> County Government | <input type="checkbox"/> Nonprofit without 501(c)3 IRS Status (Other than Institution of Higher Education) |
| <input type="checkbox"/> City or Township Government | <input type="checkbox"/> Private Institution of Higher Education |
| <input type="checkbox"/> Special District Government | <input type="checkbox"/> Individual |
| <input type="checkbox"/> Regional Organization | <input type="checkbox"/> For-Profit Organization (Other than Small Business) |
| <input type="checkbox"/> U.S. Territory or Possession | <input type="checkbox"/> Small Business |
| <input type="checkbox"/> Independent School District | <input type="checkbox"/> Hispanic-serving Institution |
| <input type="checkbox"/> Public/State Controlled Institution of Higher Learning | <input type="checkbox"/> Historically Black Colleges and Universities (HBCU's) |
| <input type="checkbox"/> Indian/Native American Tribal Government (Federally Recognized) | <input type="checkbox"/> Tribally Controlled Colleges and Universities (TCCUs) |
| <input type="checkbox"/> Indian/Native American Tribal Government (Other than Federally Recognized) | <input type="checkbox"/> Alaska Native and Native Hawaiian Serving Institutions |
| <input type="checkbox"/> Indian/Native American Tribally Designated Organization | <input type="checkbox"/> Nondomestic (non-U.S.) Entity |
| <input type="checkbox"/> Public/Indian Housing Authority | <input type="checkbox"/> Other (specify) |

7. What is the partner organization's mission? [500 characters] The Missouri History Museum seeks to deepen the understanding of past choices, present circumstances, and future possibilities; to strengthen the bonds of community; and to facilitate solutions to common problems.

8. Describe the partner organization's service area (audience served, including size, demographic characteristics and geographic area) [500 characters] At the Missouri History Museum, our audience is diverse---from young children to senior citizens of all races. While our visitors come from all over the nation, the vast majority of them are local residents from the St. Louis area. In 2007, 244,000 people visited the museum and in 2008, our audience was more than 290,000.

9. List the partner's key roles and responsibilities in the project: [1000 characters] Commit staff time and expertise in historic documents, digitization and outreach; enhance MHM staff's knowledge of digitization workflow and encoding systems. MHM personnel time: Assistant Librarian Jason Stratman, 5%, Year 1 of project; Librarian Emily Jaycox, 5%, Years 1-2.

Make MHM historic documents available to project.

a. St. Louis city directories published between 1821 and 1869: Make available for off site imaging. Washington University will provide to MHM copies of the newly imaged works with related OCR and TEI files for use on its website and other institutional projects.

b. Additional manumission related documents from MHM archives: Commit to imaging in house.

c. Previously imaged, relevant works: make available to this project for enhanced markup and access. Specifically, four city directories from the 1860's; approximately 1100 Free Negro Bonds; relevant documents from the Slavery Collection; and The Colored Aristocracy of St. Louis by Cyprian Clamorgan, 1858.

Please note:

A. Submission of this application by the Authorized Representative of the applicant organization reflects the partner organization's agreement with the following statements:

- We will carry out the activities described above and in the application narrative.
- We will use any federal funds we receive from the applicant organization in accordance with applicable federal laws and regulations as set forth in the program guidelines and the terms and conditions of the grant award.
- We assure that our facilities and programs comply with the applicable federal requirements and laws as set forth in the program guidelines.

B. Prior to submission of the application, the applicant will ensure that the partner organization has provided to the applicant a signed original of this Partnership Statement for the applicant's records. Such original will be made available to IMLS, if requested by IMLS.

SAMPLE