

**Transcription for Improved Research, Teaching and Learning at Harvard:
Final Project Report, November 16, 2012**

Project summary

The goal of this project was to identify an online transcription tool for use at Harvard from the many candidates already available. The project was a collaborative effort by the Schlesinger Library (Amy Benson), Houghton Library (Susan Pyzynski), and the Harvard University Archives (Robin McElheny).

In the project's first phase (6/1-11/15/2012), the project team assessed the functionality and platforms of existing tools, preliminarily assessed user needs, selected tools for testing, and identified end-user test participants. In the second phase, if continuation of the project is funded, (12/1/2012-4/30/2013), the team will conduct a test of selected tools and will evaluate the tools' success in terms of adoption by specific research communities and integration into classrooms, with a final report and recommendations for implementation at Harvard.

Phase 1 activities, June 1 to November 15, 2012

In addition to meeting regularly for the purposes of planning and review, we accomplished the following:

1. Surveyed a potential audience for the transcription tool
A survey of Harvard faculty, students, and other scholars was conducted in August 2012. Survey responses were enthusiastic and supportive, with 33% willing to participate in the eventual testing of transcription tools. A primary impetus for this project was the great quantity of handwritten documents that have been digitized by the Harvard Library, most of which are not amenable to OCR. Responses from faculty revealed that they also have documents and electronic content that they would like to be able to transcribe, or have students transcribe in their classes.
2. Created a list of functional criteria for evaluating tools
The project team also developed a list of functionalities and other criteria to be used in evaluating the implemented transcription tools. The list was based on extensive reading about existing projects and informed by comments and suggestions received in the surveys.
3. Selected four transcription tools for testing in Phase 2.
 - a. Transcribr, NARA
 - b. Scripto, George Mason University
 - c. Harold "Doc" Edgerton Notebooks Project, MIT
 - d. Islandora TEI Editor, University of Prince Edward Island
4. Consulted with colleagues working on related projects
 - a. In July and August, had several conversations with Fran O'Donnell and Cliff Wunderlich regarding their Library Lab project for annotating finding aids.
 - b. On 7/27, held a conference call with Suzanna Lisanti, Project Manager for the Edgerton Project at MIT, who is willing to share code for her project's transcription tool.
 - c. On 8/29, met with Robin Wendler from HUIT LTS and Andrea Goethals and Wendy Gogel, both from Harvard Library PCDI, to discuss linking transcriptions to digital objects in the DRS, integrating transcription metadata with digital objects, and to review suggested features or formats for transcription metadata. They endorsed the

desirability of end-user interactivity with digital content that underlies this project and believe that it will align with the architecture of DRS2. (Meeting, 8/29/2012)

- d. On 9/10, met with Philip Desenne, with HUIT Academic Computing, regarding his Annotation API project and possible overlap/collaboration.
 - e. Met with Sebastian Diaz on 9/18 to discuss implementation of the four transcription tools for testing, delivery of test images, and plan timeline.
5. Participated in Library Lab outreach activities
 - a. Library Lab Lightning Round, July 26, 2012
 - b. Library Lab Showcase, November 14, 2012
 6. Selected digital content for test transcriptions
As of October 15, each project repository submitted a set of digitized documents (20-25 images/repository = 60 total) for preliminary testing of the implemented tools.

Challenges – please explain anything you couldn't do

A major aim of this project was to try out test implementations of several existing tools using sample content generated by the participating libraries. Each library sent sample content, but none of the existing tools have yet been installed, so no testing could occur. Tools were to be implemented by October 15, 2012, but that didn't happen. A November 1 check-in with the programmer assigned to the project received no reply.

Next steps

The project team would like to continue to work toward our initial goals. Attached is an appendix showing a proposed timeline and budget for completing the work of the project.

Phase 1 budget

With the exception of a one-hour meeting with Sebastian Diaz on October 9, 2012, no money was spent on the project.

Request for Continued Support for *Transcription for Improved Research, Teaching and Learning at Harvard Library Lab*

The project team requests continued Library Lab support to complete the work of the second phase of the project. Continued support would require additional funding and would follow the timeline described below.

Summary

Phase 2 of the project would involve getting the selected transcription tools test installations up and running and testing them using a diversity of material from the three repositories. Evaluations of the tools will be performed by library and archives staff, student and faculty volunteers, as well as a hired student assistant. After the test period, the team will evaluate the tools' success in terms of survey responses, usefulness to specific research communities, and for integration into classrooms, concluding with a final report and recommendations for implementation at Harvard.

Estimated time required: 5 months, December 2012-April 2013, see timeline below for details
Estimated total budget: \$8,200

Budget and resource requirements detail:

1. Programmer time and effort to implement the transcription tools for testing. Th amount of time required to complete this task is unknown, but possibly 8 hours per tool @ \$100/hour. ~\$3,200
 - a. Implementation of four transcription tools:
 - i. Transcribr, NARA
 - ii. Scripto, George Mason University
 - iii. Harold "Doc" Edgerton Notebooks Project, MIT
 - iv. Islandora TEI Editor, University of Prince Edward Island
2. Adapt and additionally configure the transcription tools in response to testing and feedback.
 - a. Very difficult to estimate the amount of programming time this will require. Perhaps another 10 hours of work @ \$100/hour: ~\$1,000.
3. Undertake actual transcriptions, test and evaluate the possible tools.
 - a. One student project assistant for transcription work across all three libraries, estimated at \$4,000 (10 hours per week for 20 weeks at \$20 per hour)
 - b. If possible to do so, we may request funds for additional transcription work to provide more robust testing and results.

At the end of the pilot project, archivists and librarians will assess the following aspects of the test system through user studies and statistical reports, possibly partnering with the Simmons GSLIS:

- Platform on which the tool is built
- Ease of use
- Classroom adoption and feedback
- Popularity of tool
- Reasonable accuracy and utility of resulting transcriptions
- Ease of system management
- Success, or reasonable expectation of success with integrating the tool into Harvard's existing systems
- Researcher interest in transcribed documents

Detailed Timeline, December 2012-April 2013

- Transcription Tools implemented for testing by December 7, 2012.
 - Transcribr, NARA
 - Scripto, George Mason University
 - Harold Edgerton Project, MIT
 - Islandora TEI Editor, University of Prince Edward Island
- Large set of test images to Berkman programmer by December 7, 2012.
- Prepare test documentation/instructions by December 21, 2012
- Test period for staff, January 2012. Additionally in January:
 - Update documentation/instructions as needed.
 - Develop and test feedback survey form.
 - Hire student assistant for assessment and survey during Jan/Feb test session, and for reporting
- Two-month test period for faculty and students, February-March 2013
 - Faculty, students, scholars and other complete testing of tools and evaluation/survey.
- Project reporting, March-April 2013
 - Compile survey statistics and comments
 - Submit final report and recommendations by April 30, 2013