Contribution ID: **37**   Type: **not specified**

Publishing transcriptions as annotations on source images

Wednesday, 22 June 2011 13:50 (30 minutes)

T-PEN is a transcription tool that allows users to easily transcribe from digitized images of unpublished manuscripts. These images are made available from a growing number of large document repositories. T-PEN uses line segmentation to present the transcriber with one line at a time to transcribe, and preserves the association between the line and the portion of the image it transcribes. One of the ways users can output their transcriptions is as a set of OAC annotations on the original source images.

My talk will discuss the three ways in which OAC publication can enhance the user experience at the document repository. First, having a full or partial transcription can greatly enhance the searchability of a repository. Most repositories currently only search metadata and document catalogs, even when a transcription is available in some digital form. Second, the transcription can be displayed in the repository's image viewing environment in a number of ways that readers may find useful. The line by line alignment allows the UI to reveal only a single line at a time, or a fully visible and aligned transcript. Finally, this means of publication connects the transcribed text to the original images, and thus the original document, in a way that is permanent but also open, inviting use of the information in new and unexpected ways.

Additionally, I will discuss a number of use cases in which the transcriber is using the published transcription annotations to explore complex relationships within their transcribed text.

Presenter: Mr DEERING, Jon (Saint-Louis University (USA))

Session Classification: Plenary 1 : Towards Machine-Actionable Scholarly Communication