IMLS Grant Conference Call—October 12, 2005

1. Greg Crane verified that Alison Jones has all of the information needed for Tufts' portion of the semi-annual report to IMLS. She will send it to Jim Rettig before the end of this week.

2. Andrew Rouner explained that a preliminary splash Web site exists but has not been made live pending release and implementation of release 12 of DLXS. Its release date has been delayed repeatedly since August. The most recent date cited for release is October 21. We are currently running release 11a.

   January is our target date for making the Web site live.

3. Rachel Frick and Andrew Rouner reported on the status of each of the three newspapers:
   - Richmond Dispatch—We are nearly finished checking everything received. This week we will send the last of files that need to be returned to Digital Divide for further work. We expect all of these to be back to us by the end of this month.
   - Liberator—We have images and clean OCR
   - Ledger—Some images include two pages; these need to be redone. We have images and dirty OCR. We will get clean OCR for April and May 1865 (when the Richmond Dispatch was not publishing) and for the period just before through just after the battle of Gettysburg.

   The quality of OCRR and the accuracy of text after clean-up depends in part on the quality of the image scanned.

4. It is time to develop contextual information for the Web site. Bob Kenzer and Jim Gwin will develop this information. Rachel Frick will talk with Alison Jones about developing a how-to explanation for ways to search the site. We need to include prominent access to the user survey. Bob Kenzer will provide links to relevant Civil War Web sites.

5. Greg Crane reported on developments at Perseus:
   - They have been working on developing tools to search personal, corporate, and place names so that these tools can differentiate between these types of names and other instances of the same word or among various individuals with similar names.
   - They are trying to finish the baseline write-up on how well these indexing/searching tools work. There is a difference in context in which a word appears depending upon whether it is in the narrative in a book or in a newspaper article.
   - They are studying what sort of results different levels of investment yield— e.g., how much more does clean OCR cost compared to dirty OCR and is the difference worth the cost?
   - Perseus has *Harper's Gazetteer of the World* marked up in TEI. However TEI does not have the capability to turn this information into a database. Alison is working on this project and will document it.

Summary drafted by James Rettig

- Perseus and UR are putting things into Fedora in different ways. It will be worthwhile to compare these.