

CS-577 Project Proposal

Sponsor: Barbara Shepard <bshepard@usc.edu>

Partners: Phil Ethington, Wayne Shoaf, Marianne Afifi, deLA Technical Advisory Group

Background

Phil Ethington has jointly applied for an NEH grant to create a Web encyclopedia of Los Angeles with UCLA – deLA. It is likely that we will get the grant and the project would begin in July 2003, continuing on for several years. The creation of content for the encyclopedia is expected to begin by 2005.

Problem to solve

The content of the encyclopedia is planned to be delivered via the CIS but since the CIS is really a metadata and digital asset system, the LA encyclopedia requires additional functionality for full text encoding and access. The project would be to explore the different text encoding (TEI) standards, determine the tools, if any, available for each, identify the requirements for the LA encyclopedia and plan out cost, strategy, and expertise required to encode the encyclopedia and get it into the CIS system and its Oracle repository. Determine if the text encoded elements should be incorporated into the metadata structure and Oracle table or if they should be in a separate table with a separate structure. Determine if the keyword index should include the full text encyclopedia entries or if we should use a separate search option for full text encyclopedia searching. Identify the elements that should be encoded in the encyclopedia such as references and biographies and develop a strategy to handle the different encoded elements. Determine if a separate search engine is required for the full text searching vs. metadata searching and if yes, identify which search engines would be appropriate for this project. Investigate other full text resources available on the Web to use as models for developing our requirements.