The Walt Whitman Archive: Archivist-Scholar Collaboration in Description and Representation

Kenneth Price (kprice2@unl.edu)

University of Nebraska-Lincoln

Katherine Walter (kwalter1@unl.edu)

University of Nebraska-Lincoln

Terence Catapano (thc4@columbia.edu)

Columbia University

Daniel Pitti (dpitti@virginia.edu)

University of Virginia

Topic

Archivists, librarians, scholars, and technologists are collaborating to build *The Walt Whitman Archive*, an emerging digital thematic research collection that sets out to make Whitman's vast work easily and conveniently accessible to students, researchers, and the general public. This massive undertaking is complicated in part by the physical dispersal of Whitman's manuscripts, which are located in more than seventy repositories in the United States, the United Kingdom, and in France. The sheer volume of materials produced by Whitman has required the current project team to narrow its focus, initially at least, to a more manageable subset of the manuscripts - namely, the poetry manuscripts.

While network and computer technologies have made it possible to build a virtual archive, the intellectual and technical complexities in creating and maintaining the collection in accordance with archival and scholarly standards require an unusual and close collaboration across professional communities and among multiple institutions. Several institutions are intimately involved in the project including the University of Nebraska-Lincoln, the Institute for Advanced Technology in the Humanities at University of Virginia, the New York Public Library, the Harry Ransom Center at the University of Texas at Austin, and Duke University. The project as a whole has contributed to a clearer understanding of technical issues relating to standards that are still in development, and especially to the integration of those standards.

The poetry manuscripts, which have been the most recent focus of the Whitman project team, are scattered in over thirty repositories. At the 2002 ALLC/ACH conference in Tübingen, Germany, a preliminary report on the project entitled *Ordering Chaos: A Virtual Archive of Whitman s Manuscripts* was presented by Mary Ellen Ducey, Andrew Jewell, and Kenneth M. Price. Subsequently, the Whitman EAD project team has successfully created *An Integrated Guide to Walt Whitman's Poetry Manuscripts* using Encoded Archival Description (EAD). XSLT stylesheets are used to harvest information from various repositories' finding aids so as to create an integrated finding aid with links back to the original versions. As the project comes to closure, participants have found it both exciting and revealing on many levels. Representing the different community perspectives (scholar, archivist, librarian, and technologist), the speakers will explore both the opportunities and the challenges of working together and will discuss the implications of such collaboration for the future of each profession.

Organization

Kenneth M. Price, Co-Director of the *Walt Whitman Archive* and recently named Co-Director of Digital Research in the Humanities at the University of Nebraska-Lincoln, will describe the widely distributed Whitman manuscripts, the complex history and publication of Whitman, and the history and objectives of the *Archive*. He will discuss the reasons behind the decision to use item-level EAD as a means of bibliographic and editorial control and of user access. Our project is demonstrating the power of EAD to pull together dispersed collections and create a single, scholarly-oriented view or collocation of the materials. We are also addressing an unresolved issue in digital scholarship, namely how best to integrate description and transcription (EAD and Text Encoding Initiative [TEI] files).

Katherine L. Walter, Chair of Digital Initiatives & Special Collections and, with Price, Co-Director of Digital Research in the Humanities at the University of Nebraska-Lincoln, will describe collaborative efforts to provide integrated descriptive access to Walt Whitman's poetry manuscripts. In some cases, our EAD files are based upon encoding previously done by the holding repositories themselves; in other cases, we have created EAD files based upon paper records. We invariably add scholarly information to records, and we offer this additional information back to the individual repositories. Whitman scholarship is complicated by the fact that the poet only occasionally titled his manuscripts, and when he did, he often used a title different from that employed in any of the six distinct editions of *Leaves of Grass*. The project is ordinarily able to identify manuscripts that puzzle non-specialists, and we also supply date range, uniform title, and Whitman work IDs within the files.

Terence Catapano, Librarian at Columbia University, will discuss the complementary use of current technical standards, in particular EAD, EAC, METS, MODS, and TEI. It is still an open question how these overlapping standards - created by various communities - can be best integrated and used effectively in this kind of highly detailed collection. The *Whitman Archive* is sufficiently large, ambitious, and visible to make it a good case study for testing the integration of metadata standards. We have made significant progress in the use of TEI and EAD; we have recently begun to employ METS in relation to EAD and TEI; we plan soon to work on METS in relation to MODS as well. The periodical printings of Whitman's poetry will be used to research the use of MODS and its integration with the other metadata standards. One challenge is to figure out what role each standard is to have, and how they are to interrelate. For example, descriptive metadata resides in EAD (its primary purpose), in TEI headers (a secondary purpose), and in METS. Which of the three has the authoritative data, and which data should be derived from this authoritative source?

Daniel Pitti, Associate Director of the Institute for Advanced Technology in the Humanities of the University of Virginia, will moderate the session and, in conclusion, will discuss the implications of collaboration for the future of digital scholarship. Digital thematic research collections are valuable resources being developed through collaborations between the library/archival and scholarly communities. One of the points that has been made regarding such collections is that the essential standards shaping the infrastructure for these collections are being developed primarily in the library/archival communities. *The Walt Whitman Archive* is demonstrating that a strong collaboration with equally important contributions from different professional communities working together offers another important model for the future of digital scholarship, and for the development of standards.