1. Cover Page

Award Number: **LG-06-05-0101-05**

Awardee Institution Name: **University of Nebraska-Lincoln**

Final Narrative Report

Period Covered by the Report: **1 Nov 2005 through 30 April 2008**

Project Director Name and Title: **Katherine L. Walter**, Co-principal investigator; Chair, Digital Initiatives & Special Collections; and Co-Director, Center for Digital Research in the Humanities

Telephone: **(402) 472-3939**

Email: **kwalter1@unl.edu**

2. Description:

What is the Purpose of the Project?
The University of Nebraska-Lincoln (UNL) Libraries coordinated a demonstration project, titled *Interoperability of Metadata for Thematic Research Collections: A Model Based on the Walt Whitman Archive,* or *IOM*, to create a model Metadata Encoding and Transmission Standard (METS) Profile for digital thematic research collections, using the *Walt Whitman Archive*, http://www.whitmanarchive.org, as a test case. The *Whitman Archive* is a complex project that offered the opportunity to explore the promise of METS as a means to coordinate and encourage interoperability of metadata standards (TEI, EAD, TIFF, and MODS). UNL and the research team developed METS Profiles and METS instance records for representative objects, and tested the use of METS to manage submission and retrieval of diverse collection materials in two different digital library catalogs—one at Brown University and the other at the University of Virginia. At the beginning of this grant, no METS Profile for digital thematic research collections had been developed, and there had not been a demonstration of the effectiveness of METS as an integration tool for such collections.

The following report is made possible through the efforts of a talented and hard-working research team:

Brett Barney (Research Assistant Professor and Senior Associate Editor, *Walt Whitman Archive*, UNL)
Stacey Berry (Postdoctoral Research Associate, *Civil War Washington*, UNL)
Mary Bolin (Chair, Technical Services Department, UNL)
Terence Catapano (Special Collections Columbia University)
Julia Flanders (Director, Brown Women Writers Project and previous board president, TEI Consortium)
Daniel Pitti (Associate Director, IATH at Virginia and principal architect of EAD)
Kenneth M. Price (Co-director, *Walt Whitman Archive* and Center for Digital Research in the Humanities, UNL)
Katherine Walter (Professor and Chair, Digital Initiatives & Special Collections and co-director, Center for Digital Research in the Humanities, UNL)
Patrick Yott (Head, Digital Services, Brown University Library Center for Digital Initiatives and member of the METS Implementation Board)

What activities or services have been carried out with project funds to support the purpose of the project?

Among activities covered by project funds have been the following:

- METS Profiles relating to the *Walt Whitman Archive* have been submitted to the METS Editorial Board for inclusion in the METS Implementation Registry.

- Redundancies and overlap in the metadata developed by different communities were identified by the team at UNL. In the process, the larger team discussed the redundancies and concluded that most apparent redundancies in fact have

different functionality. Only a few redundancies could be streamlined by greater communication among standards communities.

- The University of Virginia and Brown University have successfully ingested digital objects into two open source catalogs. Selected reports relating to the ingests have been developed and links to these reports have been posted on the Interoperability of Metadata project page on the Center for Digital Research in the Humanities site at http://cdrh.unl.edu/projects/pages/interoperability_metadata. Metrics for the project are also publicly available on the CDRH site.

- Metadata is exposed for OAI harvesting. It is possible to map MODS to Dublin Core, and MODS is integral to the METS instance records created for this project.

- Several presentations relating to the project have been made. At IMLS WebWise 2006 the intent of the *Interoperability of Metadata* project was discussed as part of Katherine Walter's presentation on the IMLS-funded integrated guide to the dispersed manuscripts of Walt Whitman. Walter also moderated a very well-attended panel devoted to the *IOM* project at the international Digital Humanities 2007 conference during the first concurrent paper session. Brett Barney, Julia Flanders, Terence Catapano, and Daniel Pitti all gave excellent presentations on various aspects of our research with cogent observations on the mixed results we were receiving. The audience had many questions and observations, especially by scholars involved in thematic research. A paper was delivered by Terry Catapano about our discussions and preliminary work at the Mid-Atlantic Archives Conference in 2006, and at the December 2007 CNI meeting in Minneapolis Katherine Walter discussed the Interoperability of Metadata project as part of her presentation about the Center for Digital Research in the Humanities. Members of the project team, Catapano and Yott, will be proposing a presentation at the Digital Library Federation meeting in Providence, RI this fall.

- Several multi-institutional meetings or consultations have been held at the University of Nebraska-Lincoln to explore ways in which various metadata overlap, to discuss the design of MODS and METS instance records for thematic research collections, to develop METS Profiles that are needed for various types of objects represented in a complex thematic research collection, to create an overarching Profile for the Archive as whole, and to discuss ingestions into the open source catalogs at Brown University and the University of Virginia, as well as other related issues.

- Other meetings have occurred at conferences attended by various members of the research team, including the Digital Humanities 2007 conference and the Nov 2007 Text Encoding Initiative Conference with costs covered by the participants and their home institutions.

What are the outputs of the project activities or services to support the purpose of the project? Explain what documentation is used to report the outputs.

- A MODS template was developed by Mary Bolin and revised for use in the METS instance records for the *Walt Whitman Archive* by the project team.
- METS Profiles have been submitted to the METS Profiles list for discussion and inclusion in the METS implementation registry.
- The issue of behaviors discussed under "outcomes" in this report has been submitted to the METS Implementation Board.
- Selected reports have been posted on the *Interoperability of Metadata for Thematic Research Collections* web page, including a chart on redundancies in metadata developed by the University of Nebraska-Lincoln, a report on the ingestion at the University of Virginia, and documentation from Brown University.
- Sample submission packages have been posted in a zip file on the project web page.

What are the outcomes of the project activities or services to support the purpose of the project? Explain what documentation is used to report the outcomes.

Among planned outcomes were:
- publishing the final METS Profile in the METS registry. A METS Profile for the Walt Whitman Archives has been created and submitted to the METS Implementation list for consideration. While our hope was that the profile for the Archive could be used as a model for other digital thematic research collections that also employ multiple metadata standards, the jury on this point is still out. The METS Implementation Board has been made aware of some of the issues mentioned in the conclusions below.
- making the metrics for the ingests publicly available. Metrics are available to the public at http://cdrh.unl.edu/projects/pages/interoperability_metadata.php
- ingesting the materials into two online catalogs. Completed.
- ensuring that metadata can be harvested in OAI. Metadata created by the *Walt Whitman Archive* can be harvested. Since the project has been completed, UNL is participating in a beta test using the TEI files of the Walt Whitman Archive with Innovative Interfaces, Inc. on their Encore product.

Some conclusions based on this project:

- ➢ Development of Submission Information Packages for using METS to ingest thematic research collection objects into a digital library can be time-consuming and complicated; at the very least, it requires a good understanding of how a particular open source catalog ingests digital objects. It is thus not realistic for a scholar or his/her team to be responsible for preparing the packages for libraries to ingest. Rather, it may make more sense for a scholarly research team to expose the thematic research collection metadata so that libraries interested in capturing objects in the collection may do so.

- ➢ As noted in the Brown University report: "We discovered that there is no efficient way to associate an XSLT document with an XML file in a single METS instance, particularly when there are multiple XSLT files that relate to the XML source. This is particularly problematic as the combination of the source tree with the XSLT represents what the scholars' intent was in creating the object being shared." Patrick Yott from Brown is a member of the METS editorial board and has brought this issue to the board's attention and hopes that it will be addressed in a future revision of the METS standard. It is possible that the work that the Digital Library Federation is doing with asset actions will provide a solution.

- ➢ Also in the Brown report: "METS Profiles may become too narrative and thus not machine-actionable. Having a more actionable profile should make the transformation and validation tasks required for successful metadata interoperability much easier to design and execute. This problem is one the METS board is currently attempting to address."

- ➢ Noted in the University of Virginia report by Daniel Pitti: "The most problematic category is behaviors, that is, the computer programs that render object representations for human use or that control and process the representations in order to maintain and preserve them. There programs are intricately dependent on the data and metadata. Any changes in the expected underlying markup semantics and structure, or content formatting of the data or metadata, will more often than not require revision of the programs. Migrating data and metadata developed in one technology environment into a different technology environment in order to collect them, may require completely rewriting many or perhaps all of the programs, depending on the differences in the two environments and the objectives of the collecting environment. Many programs are based on standards—standards change. Ongoing changes in the standards will inevitably require ongoing revision and maintenance in the programs. Programs are also dependent on other programs, which may in turn be dependent on other programs. All are subject to change."

- ➢ While there is a working consensus on the types of metadata required (descriptive, technical, rights, preservation, structural, and behavioral), there is no working consensus on the details of each. Pitti notes that the least problematic is descriptive data.

- ➢ While repositories, consortia, and projects are attempting to implement standards-based digital scholarship and digital libraries that are interoperable or shareable, the choices made in the application of the standards are highly subjective—based on the implementation context, and on particular technologies chosen to achieve the objectives. There are different approaches to developing thematic digital collections—one approach might be to support in-depth research and analysis, another to provide basic reading access to digital content more or less based on object analogs of traditional media.

Report other results of the project activities.

As work progressed on this demonstration project, many questions were raised, and in fact, more questions have arisen than answers. The other major issue not previously discussed was that scholars and librarians/ archivists have different assumptions about ingestion. In her presentation at the Digital Humanities 2007 conference, Julia Flanders, Director of the Brown Women Writers Project and a dedicated member of the TEI Consortium, noted that there is a question of how thematic research collections fit into our overall understanding of the way that digital materials are created and used. These collections combine a comprehensive approach to content with a non-trivial element of scholarly shaping: they represent a perspective, an argument, a theme. This fact raises several important issues: to what extent can these intensions be discovered and reconstructed? how do these intents affect the ingestion process? if the library wishes to archive or ingest the scholarly or interpretative dimension of a collection, is that dimension represented in the data or metadata, or is it instantiated as well through interface choices and features? In the approaches to ingestion, Brown attempted to preserve the look and feel of the *Walt Whitman Archive* objects and Virginia did not. These are very different approaches, and ones that made this demonstration project more valuable.

Additional comments/anecdotal information. One of the observations Walter made as the project neared completion is that the library community has benefited from having had years of experience in dealing with MARC format, from the rule interpretations issued by the Library of Congress, and from involvement of the Cataloging and Classification Section of the Association for Library Collections and Technical Collections and other groups in discussing cataloging rules. In comparison, other standards are babes in arms. Projects like this one point out how far the different standards bodies have to go and the importance of developing rules for consistency of approach in encoding—regardless of the standard.

Catapano notes: "In this demonstration project, the *metadata* interoperability was largely accomplished; however, the *object* interoperability remains problematic. Issues of interpretation and the semantics of METS remain."

3. **Certification (to be included at the end of Performance Description)**

In submitting this report, I certify that all of the information is true and correct to the best of my knowledge.

Name and title of the person submitting the report and date of submission

_____   _____   _____
Katherine L. Walter   Project co-director   31 July 2008