

## Appendix 3

### *Research Methods*

---

#### **IMLS Digital Collections and Content Project University of Illinois at Urbana Champaign**

**September 30, 2007**

During the course of the IMLS DCC project we have assembled a rich body of data using several techniques, including interviews, surveys, focus groups, content and transaction log analysis, usability studies, etc. Below we give a detailed overview of the methods design:

Stage 1, background: The first stage of data collection was a content analysis of original grant proposals to identify the kinds of institutions, range of collections, metadata schemes and standards, aims of the projects, and personnel involved across Registry participants. In cases where information was not specified in the proposal, data were collected from project web sites. These data were gathered for ninety-five digital collection projects launched between 1998 and 2002. Later, twenty-seven additional projects launched in 2003 were integrated into the analysis, resulting in a working sample of 122 projects.

Stage 2, baseline for NLG-funded collections: In 2003 two surveys were administered to the 122 Registry participants to gather data on metadata practices and perceptions. The response rates were 76% for Survey 1 and 72% for Survey 2. Survey 1 was designed to verify, update, and augment key administrative, content, and technical information gathered in the content analysis, and to identify changes in these factors since the project was first proposed. This survey asked for information about the type of material in the digital collection, what metadata schemes were used, the intended audience, and other specifics about the digital collection and its technical implementation. Survey 2 was more subjective in nature, soliciting information about how the institution would use a Registry of digital collections, what elements should be included in a collection description, and issues or problems encountered applying, using, and trying to share metadata. With that baseline in place, we are currently conducting a follow-up survey to monitor progress and change. Supplemental baseline data were collected during this stage through an e-mail follow-up survey and a focus group. These questions probed for further information about differentiation of subcollections and efficacy of metadata schemes. A focus group was conducted in March of 2004 with 11 participants representing 9 institutions to further investigate current functionality and use of the digital collections by end users.

Stage 3, case development: During the interim between the two survey rounds we studied selected projects in more depth, drawing a purposeful sample of twenty projects from the academic, public, and state libraries, museums, historical societies, and other cultural heritage institutions, to capture the full range of operations and requirements of various services and user groups represented across the Registry participants. This stage of data collection began with 23 phone interviews with resource developers to discuss their experiences with collection building and metadata application. Interviews probed for details on: 1) the history and background of the project, 2) views on collection level descriptive metadata and what is important in representing particular collections, 3) roles of a Registry, 4) staffing and technical issues or problems, and 5) sense of the usage of the digital collection. In February of 2005 a second focus group was conducted with nine resource developers to collect data on functionality and interoperability of metadata schemas used, and expectations for the project.

Stage 4, collection focus: Data collection in this stage began once the Registry went live in early 2005. The Registry was populated with collection records developed by the registry development team. Resource developers were responsible for editing the entry for their collection and for adding child or subcollection records for the main collection record. We have been tracking changes made to collection records by resource developers and conducting phone interviews to validate our observations and learn more about collection description perceptions and practices. Eleven interviews were conducted in Fall 2005. The sample was selected to represent diversity in terms of institutional affiliation, metadata schemes used, and the types of materials included in digital collections. The open-ended interview questions probed for information about the reasons for making changes or additions to collection level description, level of satisfaction with the controlled vocabulary used for subject description, ability to capture collection uniqueness in collection description, and adequacy of harvested item-level metadata for repository end-users. Six additional face-to-face interviews were conducted in February 2006 during the WebWise conference in Los Angeles, California; a follow-up interview was conducted with one of 2006 respondents at the WebWise conference in March 2007 in Washington, DC. In early 2006, another survey was distributed to the respondents of the original 2003 survey to trace changes in metadata practices over time. This follow-up round generally covered the same questions as the Survey 1 to identify changes in types of materials, metadata applications, intended audience, and other aspects of technical implementation. The response rate for this survey round was 72%.

Stage 5, baseline for LSTA-funded collections, began in Spring 2007 when we started selecting a sample of states active in LSTA-funded digitization programs and contacting the state librarians or LSTA digitization program coordinators from this sample. Nine in-depth phone interviews have been conducted with the purpose of gathering landscape information on states emphasizing or including digitization component in LSTA programs, readiness of this community to play in collection registry and item-level repository, and the best way to add collection information and/or item-level records. Respondents were asked wide range of questions that provided data comparable to previously collected from digital resource developers of NLG-funded collections: collection development policies, topic areas and types of materials emphasized in digital collections, audiences targeted by collection developers, sustainability of digital

collections, guidelines and training provided to individual projects, decisions regarding aggregation of collections and providing access to them through federated search, successes and problems in metadata selection, application, and sharing, compliance to Open Archive Initiative Protocol for Metadata Sharing, awareness of Digital Collections and Content Registry and Repository, and possible uses for a central database of IMLS- and LSTA-funded digital collections.

A series of usability tests were conducted with Registry users over the course of the project. Additionally, the Registry transaction logs were analyzed for types of searches conducted and semantic similarity between user keyword queries and terms in controlled vocabularies. A separate analysis of subject strengths of collections in the Registry and item type distribution was also performed in early 2007.

This staged and multi-method approach has allowed us to perform analysis across a large sample of projects to address a range of research questions, and to assess views and actions of digital resource developers prior to implementation of the schema for the Collection Registry and responses to the functioning prototype of the Registry and metadata repository. Data collection and analysis is ongoing through 2010.