B. Collections of Information Employing Statistical Methods

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

Opening History is a resource of national scope meant to appeal to people interested in using the kinds of historical materials held in libraries. We intend to survey 1226 reference service providers at academic and public libraries of varying sizes from throughout the U.S. Library reference service providers are required to have extensive understanding of their user communities, since they must build collections that meet their information needs. They also provide direct assistance to these user groups to assist them in finding and using information resources that meet these needs. They are the key intermediary between the public and the Opening History resource, and can speak with authority on its value to their constituencies and to them as providers of information services to these groups. The survey population will be built from two sources: (1) a listing, based on the IMLS mailing list of public libraries, used for the Public Library Survey, merged with the Center for Informatics Research in Science and Scholarship mailing list of public libraries (a more accurate listing of public libraries in the U.S. than either list alone); and (2) an NCES listing of academic libraries from the most recent survey year, 2008. This data collection purposefully excludes institutions from the sample that have contributed content to the aggregation. We will randomly sample within each of these two frames of reference. Within each frame of reference, size (given as population served for public libraries and full-time enrollment for academic libraries, standard size categorizations for these two kinds of library institutions) will represent a data point. One individual responsible for reference services will be targeted at each institution to represent that institution. Given that the survey comprises 5-point Likert scale (categorical) and open-ended questions, we are willing to accept a margin of error of 10% as sufficient to distinguish the overall degree of agreement among respondents. A power analysis to calculate minimum sample size is provided below (using margin of error 10%; confidence 95%; power .95).

Further information provided on the sample design.
The information requested in this comment was included in the original submission. The table, below, has been re-labeled to make this clearer. However, we have revised our sample sizes and expected response according to the results of a revised power analysis using different software, as explained in the response to the next OMB comment (4). Therefore, numbers in the table below have changed since the original submission (they have been reduced).

As stated in the submission to OMB, within each frame of reference (academic and public), size (given as population served for public libraries and full-time enrollment for academic libraries, standard size categorizations for these two kinds of library institutions) will represent a data point. We are not stratifying by size, but rather using two frames of reference.

Institution Type	Universe Total population	Sample Size Sample population	Expected response (50%)***
Public libraries*	8,870	540	270
Academic libraries**	3,827	540	270

This collection is not part of an IMLS program evaluation. The aim of this research is to determine how service providers in libraries perceive the quality and scope of Opening History in respect to the needs of their local user communities, with the ultimate objective of developing and improving Opening History as a resource for history researchers and the general public. The data collection is also intended to help advance the current base of knowledge and practice for digital resource developers and service providers. Therefore, the information from the survey will primarily be used to provide feedback on Opening History and to assess directions for future development, in order to continue to meet the needs of Opening History's target audiences.

2. Describe the procedures for the collection of information including:
* Statistical methodology for stratification and sample selection,
* Estimation procedure,
* Degree of accuracy needed for the purpose described in the justification,
* Unusual problems requiring specialized sampling procedures, and
* Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

We propose to use a random sampling method within each frame: academic libraries and public libraries. The survey instrument is designed to help us determine: (1) the primary user groups for Opening History associated with each institution; (2) how Opening History compares to other comparable digital resources made by the institution, in terms of both scope of coverage of history materials and the perceived quality of the resource; (3) the effectiveness of Opening History's collection-level description; and (4) suggested improvements to increase Opening History's usefulness for the library's user communities. The two frames by library type are required to assess if there are important differences between public and academic libraries of varying sizes that need to be considered in the next stages of development.

We are asking for subjective responses; the degree of accuracy is flexible, though we do want to be able to draw firm conclusions to guide future development. With the expected response rates shown in the table above, the level of precision for each population is 10%. Given the subjective nature of the questions, this will allow sufficient generalization of the evaluative responses for our purpose of guiding general directions in development and identifying perceived weaknesses in the content or presentation of the content.

A comment box for open-ended comments is offered for most questions to elicit unanticipated responses and to allow respondents to offer qualifying and contextual information. This option is important to fully take advantage of the respondents' expert professional knowledge and perspectives. For the same reason, we include open-ended questions but have limited the number to 5 (out of 16) in order to have minimal

burden on respondents. The general feedback from these questions will provide important complementary information that will elicit specific suggestions to improve usefulness of the resource, and to help place the closed-ended results in the context of the perspective of the responding institution.

Clarification on the results of the power analysis.
Several power analyses were performed to determine the minimum required sample sizes to answer potential questions. Of the tests run, the highest required sample size was used for the final required expected sample size. Power analyses were conducted using G*Power 3.1.3 (http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/). When analyzing whether a given value within a single frame a reference might vary from a reference point such as neutrality, an analysis of t-test on means differing from a constant was used. A minimum sample size of only 54 was required for a medium effect size of 0.5 with α error probability at 0.05 and Power at 0.9 in a two-tailed test. When analyzing across frames of reference to question whether two independent samples very on a given variable, an analysis of t-test on differences between two independent means was used. A minimum total sample size of 128 (64 per frame) would be required for a medium effect size of 0.5 with α error probability at 0.05 and Power at 0.9. When analyzing a question such as whether a grouped variable affected a mean, a fixed effects Anova power analysis was used with a maximum of 10 groupings considered to coincide with the user categories. A minimum sample size of 270 was required for a medium effect size of 0.25 with α error probability at 0.05 and Power at 0.8. Therefore, the minimum sample size was set to 270 based on this calculation.

Clarification on proposing division of academic libraries to include college vs. K-12.
K-12 libraries are not included in this survey, as they do not represent a primary user group for Opening History. "Academic libraries" is widely understood within the field to refer to libraries at post-secondary institutions. For example, the Academic Libraries Survey (ALS) conducted by the United States Department of Education's National Center for Education Statistics (NCES) defines an academic library as "an entity in a postsecondary institution that provides all of the following: an organized collection of printed or other materials, or a combination thereof; a staff trained to provide and interpret such materials as required to meet the informational, cultural, recreational, or educational needs of the clientele; an established schedule in which services of the staff are available to the clientele; and the physical facilities necessary to support such a collection, staff, and schedule…When academic libraries are referred to in this report, they will always be entities that are informational resources within degree-granting postsecondary institutions in the United States, including institutions that are eligible for Title IV aid and branch campuses of Title IV-eligible institutions" (http://nces.ed.gov/pubs2010/2010348/index.asp). We use the same definition for this survey.

Further information on IMLS use of standard size categories.
Population served and full-time enrollments are standard size measures for public and academic libraries, respectively. Population served (i.e. population of legal service area) is a key demographic metric in the IMLS Public Libraries Survey, the most recent publication of which is available at <http://harvester.census.gov/imls/pubs/Publications/pls2009.pdf>. In addition, as discussed below, population served is used by the American Library Association and other organizations as a standard size measure for determining institutional costs of membership. The Academic Libraries Survey (ALS) conducted by the United States Department of Education's National Center for Education Statistics (NCES) reports academic library size in terms of full-time enrollment: <http://nces.ed.gov/surveys/libraries/academic.asp>.

Cutoffs for the categories of small, medium, and large for public libraries have been standardized by the American Library Association's Definition of Levels for Organizational Membership: http://www.ala.org/ala/membership/aladues/index.cfm. The NCES ALS referenced above makes use of 6

size categories for academic libraries (Less than 1,000; 1,000-2,999; 3,000-4,999; 5,000-9,999; 10,000-19,999; 20,000 or more). We have collapsed these into 3 categories (Less than 3,000; 3,000 – 9,999; and 10,000 or more) in part to decrease the navigational, selection, and time burden for the user, in part to parallel the options provided for public library responses, and because we do not needing more detailed data to meet the analysis objectives of the project.

Further information on open ended comment boxes, presentation plan and character limit.
Comments will be categorized according to a coding schema grounded on themes or patterns present in responses. In this way, the schema will not be predefined or artificially imposed on comments. The schema will be developed through an iterative process of reading and creating or refining categories to summarize and relate open-ended comments. Once comments have been categorized or coded, overarching patterns or trends will be identified and analyzed. In reporting, comments that represent identified categories and themes will be presented. We will only report comments that (1) represent modal responses, or (2) offer insights that might motivate future lines of inquiry or suggest a need for future work. In these cases, excerpted examples may be provided (with no identifying information, as the consent and privacy statements state). We also plan to provide a character limit of 1,000 characters for open-ended responses.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

We will maximize response rates by:
- Targeting individual reference service providers at each institution for initial contact, and ensuring in the initial contact that the nature of the burden is clear to potential respondents
- Relying on a web-based rather than paper survey, all on a single page
- Reducing time burden by asking only 16 questions, of which all but 5 are closed-ended
- Ensuring that initial contact makes clear the source of the survey and the benefit of participation (exposure to a new resource, potential to improve the resource for the public)
- Sending a follow-up reminder email to non-respondents 2 weeks after issuing initial contacts; sending a second wave of follow-up reminders to remaining non-respondents after one month.

The data collection will be adequate to provide a solid assessment of the content and presentation of the Opening History resource using descriptive statistics reports.

Further information on best practices for sending advance letters, frequent reminders to nonrespondents, "Last Chance" reminders, hardcopy letters/reminders and e-mails.
We will revise our initial follow-up plans to send weekly reminders and a final "Last Chance" reminder before the end of data collection. We will also plan to send the final reminder in hardcopy in addition to email.

Further information on sampling frames to evaluate potential differences between respondents and nonrespondents.
For public libraries, we have the following pertinent information to evaluate potential differences between respondents and nonrespondents: location, population served, circulation, holdings, expenditures, and revenue. For academic libraries, we have the following: location, Carnegie classification, full-time enrollment, circulation, holdings, and expenditures.

We plan to perform supplemental non-response analysis on the dimensions that relate to location and size of institution (indicated by population served for public libraries and full-time enrollment for academic libraries), and expenditures. We expect that these three dimensions will reflect possible bias related to geographic region and institutional resource demands as indicated by extent of service population and resource expenditures. Expenditures reflect actual resource allocations by institutions and are a commonly referenced indicator (http://www.ala.org/ala/professionalresources/libfactsheets/alalibraryfactsheet04.cfm). A nonresponse analysis will be performed and shared with OMB within 1.5 months of the data collection closing out period. .

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

The survey instrument will be pretested initially on 2 local reference service providers. The reference service providers will be asked to complete the survey and then to comment on any confusion, potential ambiguities, navigational problems, or other concerns about any of the directions or questions specified by the instrument. If modifications are required based on the results of the pretest, the instrument will be tested a second time on 2 additional service providers. Because the survey is not employing new techniques, it is expected that two phases of pretesting will be sufficient.

Further clarification "frequent users" and their basis.
"General adults" is based on the commonly recognized service category "Adults," which is commonly used in public library departmental divisions and referenced in library studies including the following:
- Stephens, A.K. (2006). Twenty-First Century Public Library Adult Services. Reference & User Services Quarterly. 45(3), p. 223-35, which explores the provision of adult services over the last two decades through an analysis of several studies;
- McCook, Kathleen de la Peña (1986). Adult services as reflective of the changing role of the public library. RQ 26, p. 180-7.

We specify general adults to differentiate from the options of more specialized adult groups that follow.

We presume that the confusion in this question comes from these more general options: "History enthusiasts", "Genealogists", and "Professionals." These three categories were derived directly from the results of previous interviews with resource providers, who identified these as user groups for their cultural heritage materials. These interviews, which were part of an earlier study, served several purposes for the current project. The user categories were for internal purposes of understanding the user community we were designing the portal for and to inform future stages of usability testing and research; these results were not disseminated in any way. Resource providers (IMLS National Leadership Grant grantees) from 19 institutions, mostly libraries, were interviewed on technical aspects and use of their digital collections. Interviews were conducted between June 2003-September 2004 by DCC Research Assistant, Ellen Knutson. In addition, library studies make frequent reference to genealogists and library users interested in genealogy and historical materials, if not as a formal service category than as a common user type. See, for example:
- *Library Trends* 32 (1) Summer 1983: Genealogy and Libraries <http://www.ideals.illinois.edu/handle/2142/5336/browse?type=dateissued>; and
- Tibbo, H. (2002). Primarily history: historians and the search for primary source materials. Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries.

Moreover, libraries promote their programs to "history enthusiasts" and genealogists (see for example, a Seattle Public Library notice at http://www.seattlehistory.org/visit_us/calendar.php?type=8; and http://statelibrary.ncdcr.gov/patrons/genealogists.html)

Beyond these first four options, this checklist represents divisions of users into student, teacher/administrator, and faculty/scholar categories. Students are further subdivided by educational stage (primary and middle; high school; undergraduate; and graduate). These are common divisions, which have been used the U.S. Department of Education National Center for Education Statistics. (2011). See the Digest of Education Statistics, 2010 (NCES 2011-015), Chapters 1 and 2 <http://nces.ed.gov/fastfacts/display.asp?id=65 and http://nces.ed.gov/fastfacts/display.asp?id=84> for examples of divisions into 'elementary', and 'secondary', i.e., 'Grades PreK-8' and 'Grades 9-12'. See U.S. Department of Education, National Center for Education Statistics. (2009). Digest of Education Statistics, 2008 (NCES 2009-020), Chapter 3 < http://nces.ed.gov/fastfacts/display.asp?id=98> for an example of divisions of post-secondary students into 'undergraduate' and 'graduate' categories.

With this question, we are not asking for numbers of users served from these groups. Rather, we are seeking reference librarians' general impressions of the most prominent types of users of historical materials in their service communities. We will not produce statistical generalizations about library user demographics. This question is intended to help us characterize the service providers' perspectives on their history service community in order to help us gauge the fit between the service community's orientation to history and the materials provided in the Opening History resource. Reference librarians are responsible for direct interaction with their service communities--in-person, online, and over the phone-- in answering reference questions, building their collections and providing interlibrary loan, and assisting them in the use of information resources. As a routine part of library service practice, they draw on their general knowledge of these communities to make decisions about developing their collections and services.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

IMLS DCC project team members, who are all faculty or staff at the Graduate School of Library and Information Science at the University of Illinois, Urbana-Champaign – will conduct the survey:
Carole Palmer, Principle Investigator, Professor and Director, CIRSS, phone: (217) 244-0653;
Katrina Fenlon, Project Coordinator, phone: (217) 244-2164;
Jacob Jett, Collections Coordinator, phone: (217) 244-2164; and
Virgil Varvel, Research Analyst, phone: (217) 333-1980.

Further information on e-mail addresses for ALL reference service providers and how "bounce backs" and invalid e-mail addresses will be handled.
We have email addresses for 2,215 of the 8,870 public libraries and 3,034 of the 3,827 academic libraries. We have postal mail addresses for all of them. Most email addresses are for library directors, who will be appropriate contacts for each institution, because they are most fit to distribute the survey to an appropriate potential respondent in the library. CIRSS has had good success with response rates among contacts in these directories in previous national surveys. As far as dealing with missing email contacts and "bounce backs": we will first conduct our random selection across the full population. Among the randomly selected sample population, for institutions for which we do not have email contacts or for which our email contacts are inappropriate (for the purposes of this survey), we will seek a correct and relevant email contact through the institution website, when possible, or through a phone call. As a last

resort, should we fail to acquire an email address for a member of the sample population, we will rely on postal mail to distribute the initial contact and consent form, survey instrument, and follow-up contacts.