

## **IMLS Digital Collections and Content: An Assessment of Opening History**

### A. Justification

**1. Explain the circumstances that make the collection of information necessary. Identify any legal or administrative requirements that necessitate the collection. Attach a copy of the appropriate section of each statute and regulation mandating or authorizing the collection of information.**

The Museum and Library Services Act authorizes the Institute of Museum and Library Services to support the following activities: to promote improvements in library services in all types of libraries in order to better serve the people of the United States; to facilitate access to resources and in all types of libraries for the purpose of cultivating an educated and informed citizenry; and to encourage resource sharing among all types of libraries for the purpose of achieving economical and efficient delivery of library services to the public. (20 U.S.C. § 9121)

In 2007, the University of Illinois at Urbana-Champaign (UIUC) entered into a contractual agreement with IMLS to maintain and enhance the IMLS Digital Collections and Content registry and item-level metadata repository – which were established by a 2002 National Leadership Grant award to UIUC – and to conduct research on “Next Generation Digital Federations: Adding Value through Collection Evaluation, Metadata Relations and Strategic Scaling.” The IMLS Digital Collections and Content (DCC) portal provides a unique, single point of access to NLG- and selected LSTA-funded digital collections. Based on an evaluation of the IMLS DCC collection registry in 2007, which revealed U.S. History as an emerging subject strength of the aggregation, and based on a project objective to expand the collection for targeted scholarly communities (in this case, history researchers), the DCC project established Opening History in 2008 as a parallel portal to DCC, to provide access to digital collections focused on U.S. history and culture. Opening History facilitates unified access to nationally distributed library and museum resources, and encourages resource sharing among all types of libraries and museums. The Next Generation Digital Federations research initiative, using Opening History as a test bed, supports IMLS’ statutory mission – to conduct analyses, identify national needs, and identify trends for its services – through research objectives including the objectives (1) to conduct formal evaluations of IMLS DCC and Opening History content, including reference librarians as a data source in the evaluation, and (2) to use the results to expand and enhance the collection for targeted scholarly communities. The proposed data collection, a nationally scoped survey of reference librarians to evaluate Opening History, is necessary to meet these objectives of the University of Illinois’ contractual agreement with IMLS. (“Next Generation Digital Federations: Adding Value through Collection Evaluation, Metadata Relations, and Strategic Scaling”, IMLS Grant No. LG-02-02-0281.) The data collection will help us develop and improve Opening History – now the largest aggregation of digital collections focusing on U.S. history – as an unparalleled, publicly available resource for history researchers and the general public. Finally, the data collection will help us advance the current base of knowledge and practice for digital resource developers and service providers.

**2. Indicate how, by whom, and for what purpose the information is to be used. Except for a new collection, indicate the actual use the agency has made of the information received from the current collection.**

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. Because Opening History is a resource of national scope, meant to appeal to history researchers and citizens interested in historical materials, we intend to survey 1226 service providers at libraries of varying types and sizes throughout the U.S. The survey will be a brief, web-based questionnaire with 16 questions. The questions, 11 of

which are closed-ended, are aimed at determining: (1) service providers' perceived local audience for Opening History; (2) how Opening History compares to other digital resources available in their libraries, both in scope and in the perceived quality of the resource; (3) the effectiveness of Opening History's collection-level description; and (4) suggested improvements for the content and presentation of Opening History. Individually identifiable information used in development of the survey population will be kept secure and separate from the survey responses, which will not be individually identifiable to project team members. Only project researchers will have access to the database containing survey responses and to the contact information for the survey population. The information gathered will not be used in any individually identifiable way. The results of this survey will be shared in a final report to IMLS. The results may also be reported in publications of IMLS DCC research findings on large-scale aggregations, but again, data collected will only be analyzed and reported in the aggregate.

Question 1 asks respondents to identify their institution type and, based on response to question 1, question 2 will ask respondents to indicate the size of the institution according to standard metrics (for academic libraries, size is categorized by full-time enrollment; for public libraries, size is categorized by population served). Question 3 asks respondents to identify the state in which their institution is located. These are the only questions asked about the institutions, and no questions request information about individual respondents. While contact information for the survey population will be stored, the survey itself will be administered anonymously and individually identifiable information on respondents will not be linked to responses (identifiable information on respondents will only be used to identify non-respondents for follow-up contacts). Library type and size are critical data points of the collection for assessing differences among different types of libraries. Since local and regional history materials are an important part of Opening History, question 3 is essential for assessing responses by geographic region. Opening History has thus far largely grown through interaction with statewide aggregations of digital cultural heritage collections—a strategy that has proven successful for rapid growth, and which will be continued through targeted recruitment of Library Services and Technology Act (LSTA) digitization projects, which are awarded at the state level. Responses by state will help the DCC project determine future collection development priorities (e.g. if there is high interest in a state that is currently underdeveloped in Opening History, the DCC can develop a plan to recruit content from that state.)

Questions 4 and 5 ask respondents to describe the groups in their service community that use historical materials. These questions are necessary to assess the variation in the service communities associated with the two types of libraries with interests in historical content, since the aim of Opening History is to provide materials for all kinds of history researchers, including students and citizen historians. We aim to find out, for example, if reference service providers representing K-12 students have very different perspectives than those primarily serving university-based user groups. Responses to these questions will help us gauge how Opening History can be further developed for these diverse audiences. Question 6 asks respondents to specify historical topics commonly of interest to their user communities. The broad phrasing of this question is intentional; we want to ensure that respondents identify topical interests in their own words, unaffected by controlled vocabularies that reify the existing, incomplete topical hierarchies. In addition, adding a controlled vocabulary to this question (imaginably in the form of a drop-down list) would likely increase the burden on respondents and compromise the survey's navigability, since it would need to be an extensive, hierarchical list that according to previous usability tests is not well aligned with users search terms. This question is necessary to determine potential gaps or weaknesses in topical coverage of the aggregation and new areas for further development. In addition, the results will allow us to evaluate the fit of our existing topical with the natural language user vocabulary. Question 7 asks respondents to name digital history resources they make available to users; this question will help us interpret responses to following comparative questions.

Questions 8-9 ask respondents to gauge the usefulness and overall quality of Opening History compared to other digital history resources made available through their library. For 8, there are too many

possibilities to specify all possible resources for comparisons and no way to determine in advance what resources these institutions provide to their users, other than through question 7. Moreover, specifying titles would risk confusing respondents with unfamiliar resources. For 9, the question is not overly specific about what “usefulness” or “quality” mean because it could vary widely by type of user. The aim is to gauge perceived relative quality in their local setting while not confusing respondents with names of unfamiliar resources or biasing responses with names of particular resources that may or may not be relevant to a particular user group or type of library. The spaces for open-ended comments will allow respondents to clarify or add context to their responses to assist us in understanding their baseline for comparison. Question 10 is a non-comparative question about the usefulness of browsing categories available on Opening History’s homepage. Librarians are very familiar with the issues associated with browsing categories for digital resources. Their expert responses will be highly valuable in assessing the value of our current categories in relation to their users’ interests.

Questions 11-12 ask respondents to comment on Opening History, based on simple interactions with the resource. Question 11 asks respondents to search for a topic of interest. This type of question has been used successfully in other studies of users of similar resources and will allow reference service providers hands-on understanding of coverage of topics of interest, rather than some prescribed search that may not be relevant to their institution’s users. Correlated to institution type, this question will help us determine if there are coverage concerns that need to be explored further. Question 12 asks respondents to gauge the usefulness of a typical collection record. Every respondent looks at the same record, which was chosen by the DCC team as a standard collection record of fairly average descriptiveness for the collections in the aggregation. This control is needed since our aim is to assess the nature of the information provided by the metadata schema (elements designated for description), which is the same for every record, not idiosyncrasies of an individual description applied to the schema. With this approach we will be able to determine if the collection description schema, represented by a typical record, aligns with respondents’ expectations across the different types of institutions.

Questions 13-14 ask about the usefulness of Opening History as a research resource for users (question 13) and for them as service providers at their libraries (question 14). Questions 15-16 ask for open-ended comments on how Opening History might be improved in terms of both content and presentation. These questions allow respondents to provide general, qualitative information about Opening History as an information resource. Having worked through the previous closed-ended questions, the respondents will have a base of understanding of the resource and be able to reflect on that to provide more extensive, unconstrained evaluative responses. These questions are necessary for us to understand the perceived usefulness of the aggregation from the expert perspective of service providers after they have reviewed and interacted with the resource, and particularly to generate responses that were not possible within the constraints of the closed ended questions. The comments will be coded thematically to identify patterns and to generate suggestions that can be considered by the Opening History development team.

**Clarification on how information from the survey will be used.**

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.

**3. Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical, or other technological collection techniques or other forms of information**

**technology, e.g., permitting electronic submission of responses, and the basis for the decision for adopting this means of collection. Also describe any consideration of using information technology to reduce burden.**

The survey will be developed and published using a free and open-source web-survey application. A web-based survey, as opposed to a paper-based survey, reduces the burden on respondents by decreasing required response time and costs and by obviating the necessity of mailing the survey back.

**4. Describe efforts to identify duplication. Show specifically why any similar information already available cannot be used or modified for use for the purposes described in Item 2 above.**

The IMLS DCC Opening History aggregation is the largest digital cultural heritage aggregation in the United States. Its unique scale and scope make Opening History an exemplar of a new service model for distributed digital libraries. We are confident, therefore, that this data collection does not duplicate any previous collection, because no other project or aggregation is in the position to conduct a similarly large-scale evaluation of service providers' impressions of this kind of extensive federation, nor has anyone conducted evaluative research on Opening History in particular. Opening History is a relatively new resource, and in the time since its inception, we have not collected any new data from the public for evaluation purposes. Data we have collected in the past have been primarily from the institutions that specifically contribute to the IMLS DCC, a related resource, and cannot be repurposed for this evaluation.

**5. If the collection of information impacts small businesses or other small entities (Item 5 of OMB Form 83-I), describe any methods used to minimize burden.**

While the collection will include small non-profit organizations, no significant impact is expected.

**6. Describe the consequence to Federal program or policy activities if the collection is not conducted or is conducted less frequently, as well as any technical or legal obstacles to reducing burden.**

This is a one-time collection. The analysis will provide essential information for making progress on an important aspect of the cooperative agreement between IMLS and the University of Illinois, to determine next steps for development of Opening History. The reference service providers surveyed are a primary stakeholder group and their perceptions of the resource will be vital to identifying limitations of the coverage, description, and presentation that will to guide next steps for optimizing Opening History.

**7. Explain any special circumstances that would cause an information collection to be conducted in a manner:**

- \* requiring respondents to report information to the agency more often than quarterly;**
- \* requiring respondents to prepare a written response to a collection of information in fewer than 30 days after receipt of it;**
- \* requiring respondents to submit more than an original and two copies of any document;**
- \* requiring respondents to retain records, other than health, medical, government contract, grant-in-aid, or tax records, for more than three years;**
- \* in connection with a statistical survey, that is not designed to produce valid and reliable results that can be generalized to the universe of study;**
- \* requiring the use of a statistical data classification that has not been reviewed and approved by OMB;**
- \* that includes a pledge of confidentiality that is not supported by authority established in statute or regulation, that is not supported by disclosure and data security policies that are consistent with the pledge, or which unnecessarily impedes sharing of data with other agencies for compatible confidential use; or**

**\* requiring respondents to submit proprietary trade secrets, or other confidential information unless the agency can demonstrate that it has instituted procedures to protect the information's confidentiality to the extent permitted by law.**

Not applicable.

**8. If applicable, provide a copy and identify the date and page number of publication in the Federal Register of the agency's notice, required by 5 CFR 1320.8(d), soliciting comments on the information collection prior to submission to OMB. Summarize public comments received in response to that notice and describe actions taken by the agency in response to these comments. Specifically address comments received on cost and hour burden. Describe efforts to consult with persons outside the agency to obtain their views on the availability of data, frequency of collection, the clarity of instructions and recordkeeping, disclosure, or reporting format (if any), and on the data elements to be recorded, disclosed, or reported. Consultation with representatives of those from whom information is to be obtained or those who must compile records should occur at least once every 3 years - even if the collection of information activity is the same as in prior periods. There may be circumstances that may preclude consultation in a specific situation. These circumstances should be explained.**

60-Day Federal Register Notice, Vol. 75, No. 90; page 26283, Tuesday, May 11, 2010. No comments received.

**9. Explain any decision to provide any payment or gift to respondents, other than reenumeration of contractors or grantees.**

Not applicable.

**10. Describe any assurance of confidentiality provided to respondents and the basis for the assurance in statute, regulation, or agency policy.**

Contact information and survey responses, in raw or individually identifiable form, will only be accessible to project personnel, and will be stored on a secure server. Survey responses will not be linked to individually identifiable information about respondents, except to identify non-respondents for follow-up contact. A click-through consent screen at the start of the survey (along with the initial contact email) will clarify that the information gathered will not be disseminated in any individually identifiable way; that the results of this survey will be shared in a final report to IMLS, our funding agency; and that the results may also help inform publications of IMLS DCC research findings on large-scale aggregations. A textual prompt at the end of the survey will remind respondents to close their browsers in order to clear their responses to the survey.

**Further information on the privacy statement that will be included on the consent screen.**

This privacy statement that was included with the Consent document (ConsentIMLSDCC.docx) has been pasted again below. This privacy statement will appear as part of the survey software template at the top of the survey. Users will see this statement after they have clicked through the consent screen, and before beginning the survey.

***A Note On Privacy***

This survey is anonymous.

The record kept of your survey responses does not contain any identifying information about you unless a specific question in the survey has asked for this. If you have responded to a survey that used an identifying token to allow you to access the survey, you can rest assured that the identifying token is not kept with your responses. It is managed in a separate database, and will

only be updated to indicate that you have (or haven't) completed this survey. There is no way of matching identification tokens with survey responses in this survey.

**11. Provide additional justification for any questions of a sensitive nature, such as sexual behavior and attitudes, religious beliefs, and other matters that are commonly considered private. This justification should include the reasons why the agency considers the questions necessary, the specific uses to be made of the information, the explanation to be given to persons from whom the information is requested, and any steps to be taken to obtain their consent.**

There are no questions of a sensitive nature for this data collection.

**Clarification on individual information and raw data transfer.**

Individual information and raw data will not be transferred to IMLS. Contact information and survey responses, in raw or individually identifiable form, will only be accessible to project personnel, and will be stored on a secure server at the University of Illinois. Survey responses will not be linked to individually identifiable information about respondents. A click-through consent screen at the start of the survey (along with the initial contact email) will clarify that the information gathered will not be disseminated in any individually identifiable way. The data will be destroyed 36 months after its collection.

**12. Provide estimates of the hour burden of the collection of information. The statement should:**

**\* Indicate the number of respondents, frequency of response, annual hour burden, and an explanation of how the burden was estimated. Unless directed to do so, agencies should not conduct special surveys to obtain information on which to base hour burden estimates. Consultation with a sample (fewer than 10) of potential respondents is desirable. If the hour burden on respondents is expected to vary widely because of differences in activity, size, or complexity, show the range of estimated hour burden, and explain the reasons for the variance. Generally, estimates should not include burden hours for customary and usual business practices.**

**\* If this request for approval covers more than one form, provide separate hour burden estimates for each form and aggregate the hour burdens in Item 13 of OMB Form 83-I.**

**\* Provide estimates of annualized cost to respondents for the hour burdens for collections of information, identifying and using appropriate wage rate categories. The cost of contracting out or paying outside parties for information collection activities should not be included here. Instead, this cost should be included in Item 13.**

IMLS estimates the following burden for collection of this information (see chart below) :

Expected respondents (50% of sample population): 613

Estimated response time: 0.3 hours

Estimated total burden hours: 183.9

Estimated cost per respondent: \$8.03 (.3 hours x \$26.76 per hour<sup>1</sup>)

Estimated total burden: 0.3 hours x 613 x \$26.76 = \$4,921.16

Estimate based on: University of Illinois research team estimate

**13. Provide an estimate for the total annual cost burden to respondents or recordkeepers resulting from the collection of information. (Do not include the cost of any hour burden shown in Items 12 and 14).**

**\* The cost estimate should be split into two components: (a) a total capital and start-up cost**

---

<sup>1</sup> Source of salary estimate: Bureau of Labor Statistics' Occupational Employment and Wages, May 2009, 25-4021 Librarians, retrieved November 5, 2010 from <http://www.bls.gov/oes/2009/may/oes254021.htm>

component (annualized over its expected useful life) and (b) a total operation and maintenance and purchase of services component. The estimates should take into account costs associated with generating, maintaining, and disclosing or providing the information. Include descriptions of methods used to estimate major cost factors including system and technology acquisition, expected useful life of capital equipment, the discount rate(s), and the time period over which costs will be incurred. Capital and start-up costs include, among other items, preparations for collecting information such as purchasing computers and software; monitoring, sampling, drilling and testing equipment; and record storage facilities.

\* If cost estimates are expected to vary widely, agencies should present ranges of cost burdens and explain the reasons for the variance. The cost of purchasing or contracting out information collections services should be a part of this cost burden estimate. In developing cost burden estimates, agencies may consult with a sample of respondents (fewer than 10), utilize the 60-day pre-OMB submission public comment process and use 10/95 existing economic or regulatory impact analysis associated with the rulemaking containing the information collection, as appropriate.

\* Generally, estimates should not include purchases of equipment or services, or portions thereof, made: (1) prior to October 1, 1995, (2) to achieve regulatory compliance with requirements not associated with the information collection, (3) for reasons other than to provide information or keep records for the government, or (4) as part of customary and usual business or private practices.

Recordkeeping burden for respondents caused by this data collection: None. Participants will respond according to their impressions of the information needs of their user populations, and their impressions of the Opening History resource.

**14. Provide estimates of annualized costs to the Federal government. Also, provide a description of the method used to estimate cost, which should include quantification of hours, operational expenses (such as equipment, overhead, printing, and support staff), and any other expense that would not have been incurred without this collection of information. Agencies may also aggregate cost estimates from Items 12, 13, and 14 in a single table.**

This collection is part of IMLS Cooperative Agreement LG-02-02-0281, \$975,903 to the University of Illinois for three years. The primary costs for this collection will be the costs of survey design and administration, sampling, and data analysis.

Survey design, sampling, and survey administration (Principle investigator, Research Assistant and Project coordinator):

PI (1 summer month): \$10,622

Project Coordinator and RA: 100 hours at \$22/hour = \$2,200

Websurvey construction:

RA: 50 hours at \$22/hour = \$1,100

Data analysis: (this is not charged to IMLS, but to the institutional cost match)

Data analyst: 116 hours, \$10,000

Total cost (including institutional cost match, not charged to IMLS): \$23,922

**15. Explain the reasons for any program changes or adjustments reported in Items 13 or 14 of the OMB Form 83-I.**

Not applicable

**16. For collections of information whose results will be published, outline plans for tabulation and**

**publication. Address any complex analytical techniques that will be used. Provide the time schedule for the entire project, including beginning and ending dates of the collection of information, completion of report, publication dates, and other actions.**

The survey will be distributed, pending OMB clearance, in April 2011. The results will be disseminated through presentations at professional conferences, such as the American Society for Information Science and Technology (ASIS&T), and through publication in scholarly journals, such as the Journal of the American Society for Information Science and Technology (JASIS&T) and D-Lib. The results will also be given to IMLS as part of the standard project reporting process.

*April 2011:* Initial contact and survey distribution pending OMB clearance.

*Early May 2011:* Initial follow-up email to non-respondents, 2 weeks after initial contact.

*Late May 2011:* Final follow-up email to non-respondents, 4 weeks after initial contact.

*June 2011:* Close survey, 6 weeks after survey distribution.

*June – July 2011:* Data analysis. Due to the simplicity of the survey, 4 weeks after close of survey is expected to be sufficient for data analysis.

*August 2011 – May 2012:* Reporting and dissemination. Results will first be formally reported in interim report to IMLS, October 2011. Journal publication or conference presentation will follow, with submission to journals and conferences as late as spring 2012.

**17. If seeking approval to not display the expiration date for OMB approval of the information collection, explain the reasons that display would be inappropriate.**

Not applicable

**18. Explain each exception to the certification statement identified in Item 19, "Certification for Paperwork Reduction Act Submissions," of OMB Form 83-I.**

Not applicable