



*Heritage Health Information Survey 2014*

*User's Guide and Data File Documentation*

**February 2019**

Lisa M. Frehill  
Marisa Pelczar  
Matthew Birnbaum  
Benjamin Sweezy  
**Institute of Museum and Library Services**

Institute of Museum and Library Services  
Kathryn K. Matthew

Director

The Institute of Museum and Library Services is the primary source of federal support for the nation's libraries and museums. We advance, support, and empower America's museums, libraries, and related organizations through grant-making, research, and policy development. Our vision is a nation where museums and libraries work together to transform the lives of individuals and communities. To learn more, visit the [IMLS online](#) and follow us on [Facebook](#) and [Twitter](#).

As part of its mission, IMLS conducts policy research, analysis, and data collection to extend and improve the nation's museum, library, and information services. IMLS research activities are conducted in ongoing collaboration with state library administrative agencies; national, state, and regional library and museum organizations; and other relevant agencies and organizations. IMLS research activities are designed to provide consistent, reliable, and accurate indicators of the status and trends in library and museum services; and to report timely, useful, and highquality data to Congress, the states, other policymakers, practitioners, data users, and the general public.

#### Contact Information

Institute of Museum and Library Services  
955 L'Enfant Plaza North, SW Suite 4000  
Washington, DC 20024-2135  
202-653-IMLS (4657)

The IMLS Internet address is <http://www.imls.gov>.

This publication is available only on the Web at <http://www.imls.gov/research/>  
IMLS will provide an audio recording of this publication upon request.

For questions or comments, contact [Research@imls.gov](mailto:Research@imls.gov)

February 2019

#### **Suggested Citation:**

Frehill, L. M., Pelczar, M., Birnbaum, M., and Sweezy, B. (2019). *Heritage Health Information Survey 2014: User's Guide and Data File Documentation*. Institute of Museum and Library Services. Washington, DC.

Unless specifically noted, all information contained herein is in the public domain and may be used and reprinted without special permission. Citation of this source is required.

# TABLE OF CONTENTS

---

<b>Table of Contents</b> .....	<b>i</b>
<b>I. Introduction and Background</b> .....	<b>3</b>
<b>II. Questionnaire</b> .....	<b>4</b>
Online and Paper Survey .....	4
<b>III. Survey Universe and Sampling</b> .....	<b>6</b>
Universe .....	6
Sampling .....	6
Phone Verification of Sampled Group 3 Institutions.....	7
<b>IV. Data Collection Procedures</b> .....	<b>8</b>
Administering the Survey Questionnaire .....	8
Follow up .....	8
Final Dispositions .....	9
<b>V. Unit response rates, Sources of error, and weighting</b> .....	<b>10</b>
Unit Response Rates .....	10
Precision of Sample Estimates (Margin of Error).....	11
Sources of Error in the HHI 2014 Survey .....	11
Measurement Error .....	11
Technical or Data Handling Errors .....	12
Response Bias .....	12
Nonresponse Bias Survey Results .....	14
Weight Development Process .....	15
<b>VI. Post-Collection Data Processing and Data Considerations</b> .....	<b>17</b>
Data Review and Consistency Checks.....	17
Coding of Categorical and Numeric Variables .....	17
Coding of Open-Ended Response Variables .....	17
Response Option of “Don’t Know” .....	17
Institutional Staff for Conservation/Preservation (D7, D7a, D7b, D7c).....	18
Overall FTE Staff (G1a, G1b) .....	18
Annual Operating Budget (E1) and Annual Operating Budget for Preservation (E4).....	18
Data Imputation for Institutional FTE Staff (D11b, G1a, G1b), Annual Budgets (E1, E4), and Visitors (G2a, G2b, G2c).....	18
Digital Material Collections (F5k).....	19

<b>VII. Construction of the PUD.....</b>	<b>20</b>
Protecting Respondent Confidentiality .....	20
Data Quality: Item Response Rates .....	23
Variables Included to Facilitate Analysis .....	28

**List of Tables**

<b>Table 1. HHI 2014 Questionnaire Summary.....</b>	<b>4</b>
<b>Table 2. Estimated Size of Institution Universe in Initial Sampling Frame .....</b>	<b>6</b>
<b>Table 3. Number of Institutions in Sampling Frame, Sampling Rate, and Number Sampled by Group and Type of Institution .....</b>	<b>7</b>
<b>Table 4. Results of Pre-Administration Phone Verification for Sampled Institutions in Group 3 .....</b>	<b>7</b>
<b>Table 5. Number of Institutions Sent Questionnaires by Email and Mail .....</b>	<b>8</b>
<b>Table 6. HHI 2014 Survey Data Collection and Follow up Schedule.....</b>	<b>8</b>
<b>Table 7. Eligible Sample Based on Final Dispositions after Survey Administration by Group and Type of Institution .....</b>	<b>9</b>
<b>Table 8. Population and Sample Sizes, Number of Responding Institutions, Unweighted Response Rates, and Weighted Response Rates by Group and Self-Identified Type of Institution.....</b>	<b>10</b>
<b>Table 9. Maximum Margins of Error for Percentage Estimates Involving All Institutions, Size Groups, and Self-Identified Institution Types .....</b>	<b>11</b>
<b>Table 10. Representativeness Indexes by Type of Institution .....</b>	<b>13</b>
<b>Table 11. Nonresponse Bias Survey Respondents.....</b>	<b>14</b>
<b>Table 12. Summary of Nonresponse Bias Study Results .....</b>	<b>15</b>
<b>Table 13. Variables Recoded to Avoid Small Cells (Disclosure Risk) .....</b>	<b>20</b>
<b>Table 14. Topcoding Details, Continuous Items in PUD (Except Collections Variables) .....</b>	<b>21</b>
<b>Table 15. AAM Region Definitions .....</b>	<b>23</b>
<b>Table 16. Topcoding Details, Physical Collections Variables .....</b>	<b>24</b>
<b>Table 17. Topcoding Details, Digital Collections Variables.....</b>	<b>27</b>
<b>Table 18. Assignment of Institutions to the Variable rself_type .....</b>	<b>29</b>

## I. INTRODUCTION AND BACKGROUND

---

The Institute of Museum and Library Services (IMLS) is the primary source of federal support for the nation's libraries and museums. Its mission is to create strong libraries and museums that connect people to information and ideas. IMLS works at the national level and in coordination with state and local organizations to sustain heritage, culture, and knowledge; enhance learning and innovation; and support professional development. The care of collections items, using such methods as conservation science or temperature controls, covers a wide range of professional activities across U.S. cultural heritage organizations, and is at the heart of all cultural heritage and collecting practices.

The Heritage Health Information Survey (HHIS) 2014 was conducted to update a similar 2004 study<sup>1</sup> of the assessment of the condition of non-living collections at U.S. institutions, to measure the extent of preservation activities and initiatives, and to gather data about professional development and institutional needs related to collections preservation. Many of the questionnaire items repeated (or slightly modified) those that were included in the 2004 study that was funded via a grant by IMLS to Heritage Preservation.

While Heritage Preservation (HP) was the initial implementing organization for the HHI 2014 survey, RMC Research Corporation (RMC) designed and executed the institutional sampling plan and collected and edited the data.

---

<sup>1</sup> The 2004 study, titled the Heritage Health Index survey, was administered by Heritage Preservation and RMC Research Corporation via a grant from IMLS and with additional funds from other sources. In 2005, a summary publication titled "A Public Trust at Risk: The Heritage Health Index Report on the State of America's Collections" was completed and is available at: <http://www.conservation-us.org/docs/default-source/hhi/hhisummary.pdf?sfvrsn=2>.

## II. QUESTIONNAIRE

---

Table 1 shows a summary of the topics covered in the HHIS 2014 survey questionnaire (see Appendix A for the full questionnaire).

**Table 1. HHI 2014 Questionnaire Summary**

Survey Section	Topics
<b>A. Eligibility to Respond to the Survey</b>	Holds non-living collections held for preservation, non-profit status
<b>B. Institutional Type and Governance</b>	Self-identified type and functions, governance
<b>C. Environment</b>	Environmental controls, storage
<b>D. Preservation Activities</b>	Mission, preservation plans, assessment, emergency plan, staffing, activity functions, digital collections, preservation improvements needed, history of damage, extent of preservation activities
<b>E. Expenditures and Funding</b>	Operating budget, preservation activity funded by internal and external sources, preservation budget, grant funding
<b>F1 – F4. Collections Cataloging</b>	Cataloged collections, access to digitized collections
<b>F5. Collections and Holdings</b>	Number and condition of books and bound volumes, unbound sheets, photographic collections, moving image collections, recorded sound collections, art objects, historic and ethnographic objects, archaeological collections (individual, bulk), natural science specimens, digital material collections
<b>G. Staffing and Visitors</b>	Total paid/unpaid staff, number of visitors

### Online and Paper Survey

RMC developed an online survey with a comprehensive set of dynamic validation checks for online data entry, including:

- Consistency checks developed across questions,
- Possible out-of-range values defined for the numeric responses (e.g., operating budgets, number of units in a collection), and
- Error and warning messages programmed into the online survey to ensure data integrity.

The HHI survey pretest phase for digital collections indicated that many institutions needed to gather information for the survey from several people or departments within their institution. To accommodate this, the online survey had several unique features, such as

- Allowing for multiple online sessions of data entry,
- Built-in indicators to inform users of the section-by-section status of completion,
- Allowing respondents to print out the completed survey before final submission of the data, and

The survey engine mechanism provided all of the field-level validation requirements of the study. These validation checks included numerical data type validation for numeric data, email, and URL fields. Numeric, date, and single- and multiple-choice fields provided both range validation and comparison testing against other discrete fields in any section of the form for internal consistency

checks. Automatically calculated fields could be added to perform mathematic operations against any number of discrete fields in any section of the form. Individual fields could be declared as required, and the required flag could respond to customized skip logic against answers provided in other fields. The survey mechanism provided immediate feedback response to any validation errors on the respondent-side when a response failed an edit, and performed server-side validation checks to ensure data integrity. Customized pop-up help was available for all individual fields or sections.

RMC also designed a paper survey that matched the online survey, which lacked the validation checks, error and warning messages and other features in the online version. The paper survey was mailed to institutions that requested a mailed survey or for which an email address had not been obtained.

### III. SURVEY UNIVERSE AND SAMPLING

---

Heritage Preservation (HP) used a comprehensive process, described in this section, to identify the universe of collecting institutions for the HHI 2014. RMC Research Corporation (RMC), with collaboration from HP, developed and executed the sampling plan described in this section to obtain an adequate number of responding institutions, so that statistically accurate findings could be reported for the HHIS 2014 study at the national level and aggregated by institutional size and type.

#### Universe

The HHIS 2014 survey universe was based on extensive directories from the HHI 2004 survey (see Appendix B). HP reviewed the source lists that were used in HHI 2004 and identified lists that were currently active and those that had been updated. Many of the directories used in 2004 were duplicative of primary sources, such as the American Library Directory and the Official Museum Directory. Where needed, HP staff called and verified the collecting procedures for directories with staff at the federal agency, private company, or association responsible for the list.

RMC verified all institutions in the 2004 universe were included in the 2014 universe. The initial HHIS 2014 universe included 45,782 institutions with the responsibility of holding non-living collections in the public trust. These institutions were categorized into five types and three groups based on size, as shown in Table 2. Size groupings depended on the institutional type and took into account the size of the collection(s), annual budgets, size of population served, or number of staff, as described in more detail in Appendix C.

**Table 2. Estimated Size of Institution Universe in Initial Sampling Frame**

Group and Size	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
<b>Group 1, Large</b>	102	231	38	220	38	<b>629</b>
<b>Group 2, Medium</b>	42	395	15	260	54	<b>766</b>
<b>Group 3, Small</b>	813	14,123	4,268	23,291	1,892	<b>44,387</b>
<b>Total</b>	<b>957</b>	<b>14,749</b>	<b>4,321</b>	<b>23,771</b>	<b>1,984</b>	<b>45,782</b>

Organizational entities operating under a parent institution were accounted for by the parent institution only. For example, a museum with a library was asked to complete the survey for both its museum and library collections. Systems of collecting institutions that had central collections control and preservation practices, such as a library system within a university, were asked to complete the survey for the main library and departmental libraries. However, professional schools, such as a university's business school, medical school, or law school, or university museums and departmental collections (archaeology or sciences) were often not centralized in administration, and were included in the universe separately.

This universe did not include historic structures or living heritages, such as performing arts or living collections in institutions such as zoos, aquariums, and botanical gardens. Other exclusions from the study universe were: elementary and secondary school and two-year college libraries; branch public libraries; hospital libraries; prison libraries; county clerk office, law firm, newspaper, corporate, or engineering firm libraries; and for-profit organizations.

#### Sampling

Table 3 reports the sampling rate and number of units that were sampled within each of the 15 group by type strata. All 629 institutions in Group 1, all 766 institutions in Group 2, and all 813 Archives in Group 3 were included in the survey sample, due to their relatively small numbers, while

institutions of the other four types in Group 3 were randomly sampled, stratified by type and geographic region defined by the American Association of Museums (New England, Mid-Atlantic, Southeast, Mid-West, Mountain Plains, and West). The number of institutions sampled from each of these four types in Group 3 was based on obtaining a sufficient number of responding institutions in order to report statistically accurate findings at the national level and for each type of institution.

**Table 3. Number of Institutions in Sampling Frame, Sampling Rate, and Number Sampled by Group and Type of Institution**

Group and Size	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
<b>Group 1, Large</b>						
Universe	102	231	38	220	38	629
Number Sampled	102	231	38	220	38	629
<b>Group 2, Medium</b>						
Universe	42	395	15	260	54	766
Number Sampled	42	395	15	260	54	766
<b>Group 3, Small</b>						
Universe	813	14,123	4,268	23,291	1,892	44,387
Number Sampled	813	3,310	1,643	3,323	1,381	10,470
Sampling Rate	100%	23%	38%	14%	73%	
<b>Total Number Sampled</b>	<b>957</b>	<b>3,936</b>	<b>1,696</b>	<b>3,803</b>	<b>1,473</b>	<b>11,865</b>

### Phone Verification of Sampled Group 3 Institutions

The institutions in the Group 3 sampling frame were identified through multiple, possibly overlapping listings, which may have had unreliable information about these organizations, their collections, and staffing. To ensure all sampled institutions in Group 3 were eligible for inclusion in the study, RMC conducted a phone verification process between September 20, 2014 and October 20, 2014 using the protocol in Appendix D. All sampled Group 3 institutions were contacted by telephone to validate their eligibility for inclusion in the survey as non-profit institutions holding non-living collections. This phone verification process was also used to assess each institution's ability/preference to access the survey questionnaire online or by mail, and to identify the most appropriate contact within the institution to receive correspondence about the study.

Table 4 summarizes the results of the phone verification process for sampled Group 3 institutions. Overall, 2,714 (26%) of the 10,470 sampled institutions in Group 3 were found to be not eligible. The remaining 7,756 institutions constituted the eligible survey sample prior to survey administration. The verification process resulted in a revised universe of 32,314 institutions in Group 3 rather than the original 44,387 as reported in Table 3. Additionally, as shown in Table 4, 590 Group 3 institutions refused to participate in the survey and were treated as unit non-respondents.

**Table 4. Results of Pre-Administration Phone Verification for Sampled Institutions in Group 3**

Result	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
Number sampled	813	3,310	1,643	3,323	1,381	10,470
Not Eligible						
- Number	166	632	498	1,063	355	2,714
- Percent	(20.4%)	(19.1%)	(30.3%)	(32.0%)	(25.7%)	(25.9%)
Refusals	57	179	89	158	107	590
<b>Actual Sample</b>	<b>590</b>	<b>2,499</b>	<b>1,056</b>	<b>2,102</b>	<b>919</b>	<b>7,166</b>

## IV. DATA COLLECTION PROCEDURES

### Administering the Survey Questionnaire

A total of 8,561 eligible collecting institutions were sent survey questionnaires via an emailed online link or by mail on October 24, 2014 using the letter included with the questionnaire in Appendix A. Table 5 summarizes the number of institutions by group and institutional type that were sent the HHIS 2014 questionnaire via email or by physical mail. All 629 Group 1 and 766 Group 2 institutions and about 73% of the Group 3 institutions were sent a link to the online version of the HHIS 2014 questionnaire. RMC assigned each institution a unique password for accessing its survey to ensure data security. A paper copy of the questionnaire was mailed to the remaining 1,904 (27%) of the Group 3 institutions. The institutions were initially given a month to complete the survey.

**Table 5. Number of Institutions Sent Questionnaires by Email and Mail**

Group and Size	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
<b>Group 1, Large (Email)</b>	102	231	38	220	38	629
<b>Group 2, Medium (Email)</b>	42	395	15	260	54	766
<b>Group 3, Small</b>	590	2,499	1,056	2,102	919	7,166
<i>Email</i>	489	1,902	623	1,516	732	5,262
<i>Mail</i>	101	597	433	586	187	1,904
<b>Total Receiving Questionnaires</b>	734	3,125	1,109	2,582	1,011	8,561

### Follow up

Because of a low return rate within the first month, RMC and HP made multiple efforts to reach non-responding institutions. HP and RMC confirmed or corrected contact information for surveys that were not returned in order to send out the questionnaire a second time to the corrected email contact or physical address. HP made personal reminder calls to all Group 1 and some Group 2 institutions to encourage study participation. RMC staff contacted all Group 3 archives, historical societies, and scientific research organizations by telephone reminding them of the study invitation. Due to limited resources, RMC contacted via phone only half of the non-respondents from museums and libraries in Group 3.

Institutions requested extensions due to holiday staffing constraints, which resulted in an extension announcement by email and mail in November 2014. A final reminder was sent by email and mail on February 2, 2015, with a final deadline extension to February 13, 2015. All announcements and reminders were sent to those who had not responded or who had online surveys in progress. Table 6 summarizes the HHIS 2014 data collection schedule and the follow-up efforts made by RMC and HP.

**Table 6. HHI 2014 Survey Data Collection and Follow up Schedule**

Activity	Date	Survey Deadline Request
Initial Survey Invitation by Email/Mail	10/24/2014	11/24/2014
Reminder #1 (Email)	11/12/2014	11/24/2014
Reminder #2 and Extension Announcement (Email/Postcard)	11/19/2014	12/19/2014
Reminder #3 and Extension Announcement (Email/Postcard)	12/19/2014	2/13/2015
Follow-up Phone Calls	12/19/2014 - 2/1/2015	2/13/2015
Reminder #4 (Email/Postcard)	2/2/2015	2/13/2015

## Final Dispositions

During survey administration RMC obtained additional information that further reduced sample sizes with subsequent revisions to the universe within the 15 size group by institutional type strata. These changes were often due to the original inclusion of institutions with multiple locations, each of which received a copy of the survey. In some cases, such institutions identified themselves to RMC, which then confirmed this new information. These institutions were given the disposition of “Not Eligible Post-Admin” and were removed from the survey, since data associated with the entity would be included with the “parent” institution.

Table 7 reports the post-administration eligibility findings for sampled institutions by group and institution type. In Group 1, the number of institutions in the eligible sample was reduced to 586 after survey administration. In Group 2, the eligible sample decreased to 742 institutions. In Group 3, the eligible sample was reduced from 7,756 to 7,057. Thus, 8,385 of the 9,151 institutions that were considered eligible to participate prior to survey administration were still considered eligible after survey administration.

Additionally, after administering the survey, RMC identified institutions that should not have been in the sampling frame such as duplicate universe entities, institutions with no holdings for which they had preservation responsibility, and institutions that had ceased operation. These were also identified as “Not Eligible Post-Admin” and removed from the survey.

**Table 7. Eligible Sample Based on Final Dispositions after Survey Administration by Group and Type of Institution**

Group and Size	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
<b>Group 1, Large</b>						
Initial Eligible Sample	102	231	38	220	38	629
Not Eligible Post-Admin Number Percent	20 (19.6%)	16 (6.9%)	2 (5.3%)	4 (1.8%)	1 (2.6%)	43 (6.8%)
Final Eligible Sample	82	215	36	216	37	586
<b>Group 2, Medium</b>						
Initial Eligible Sample	42	395	15	260	54	766
Not Eligible Post-Admin Number Percent	2 (4.8%)	16 (4.0%)	0 (0%)	5 (1.9%)	1 (1.8%)	24 (3.1%)
Final Eligible Sample	40	379	15	255	53	742
<b>Group 3, Small</b>						
Initial Eligible Sample	647	2,678	1,145	2,260	1,026	7,756
Not Eligible Post-Admin Number Percent	102 (15.8%)	276 (10.3%)	77 (6.7%)	163 (7.2%)	81 (7.9%)	699 (9.0%)
Final Eligible Sample	545	2,402	1,068	2,097	945	7,057
<b>Total Sample</b>						
Initial Eligible Sample	791	3,304	1,198	2,740	1,118	9,151
Not Eligible Post-Admin Number Percent	124 (15.9%)	308 (9.3%)	79 (6.6%)	172 (6.3%)	83 (7.4%)	766 (8.4%)
Final Eligible Sample	667	2,996	1,119	2,568	1,035	8,385

## V. UNIT RESPONSE RATES, SOURCES OF ERROR, AND WEIGHTING

This section describes the number and percentage of institutions from each size group and institution type that responded to the HHIS 2014 survey along with the maximum margins of error for estimations of dichotomous items in the survey using the number of responses and the final eligible population sizes. Finally, this section describes the process by which population weights were developed based on unit response rates and responsiveness analysis. These weights permit data users to generalize to the population of U.S. collecting institutions.

### Unit Response Rates

Table 8 presents unweighted and weighted unit response rates. Overall, Group 1 institutions had the highest response rates, averaging 59% across all five types, while Group 3 institutions had the lowest at an average of 16%.

**Table 8. Population and Sample Sizes, Number of Responding Institutions, Unweighted Response Rates, and Weighted Response Rates by Group and Self-Identified Type of Institution**

Group and Size	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
<b>Group 1, Large</b>						
Population*	82	215	36	216	37	586
Respondents	64	124	24	123	11	346
Response Rate*	78.0%	57.7%	66.7%	56.9%	29.7%	59.0%
<b>Group 2, Medium</b>						
Population*	40	379	15	255	53	742
Respondents	11	103	10	107	11	242
Response Rate*	27.5%	27.2%	66.7%	42.0%	20.8%	32.6%
<b>Group 3, Small</b>						
Population	545	10,663	2,743	14,733	1,278	29,962
Sample	545	2,402	1,068	2,097	945	7,057
Respondents	163	302	187	345	129	1,126
Unwt Response Rate	29.9%	12.6%	17.5%	16.5%	13.7%	16.0%
Wt Response Rate	29.9%	12.1%	17.7%	16.4%	13.8%	15.1%
<b>Total</b>						
Population*	667	11,257	2,794	15,204	1,368	31,290
Sample	667	2,996	1,119	2,568	1,035	8,385
Respondents	238	529	221	575	151	1,714
Unwt Response Rate	35.7%	17.7%	19.7%	22.4%	14.6%	20.4%
Wt Response Rate	35.7%	17.0%	20.0%	22.3%	14.8%	16.7%

\*Since Group 1, Group 2 and archives within Group 3 were 100% sampled, unweighted response rates are the same as the weighted response rates.

### Precision of Sample Estimates (Margin of Error)

Based on the number of responses and population sizes reported in Table 8 the precision of institutional estimates for dichotomous survey items (e.g., yes/no) in the HHI 2014 survey. The maximum margin of error based on a 95% confidence interval was calculated for the overall responding sample, for each size group and institution type, and for each type within each size group using the following formula:

$$\text{Margin of Error} = 1.96 \times \left( \sqrt{\left(\frac{.25}{n}\right) \times \left(\frac{N-n}{N-1}\right)} \right),$$

where  $N$  = number of eligible institutions in the population and  $n$  = number of respondents.

Table 9 presents the maximum margins of error overall, by type, by size, and within type and size. With the lower response rates (see Table 8), and subsequent larger margins of error shown in Table 9 for the Group 2 institutions, users are advised to combine Group 1 and Group 2. Such comparisons are facilitated with a variable in the PUD named Group2R, which combines large/medium institutions in one category and small ones in a second category (see the size criteria described in Appendix C). Additionally, even combining Groups 1 and 2 for scientific collections results in a large margin of error (+/- 16%), therefore users are also advised to analyze data associated with scientific collections at the aggregate level without differentiating by size.

**Table 9. Maximum Margins of Error for Percentage Estimates Involving All Institutions, Size Groups, and Self-Identified Institution Types**

Group and Size	Archives	Libraries	Historical Societies	Museums	Scientific Collections	Total
Group 1, Large	2.7%	3.7%	6.9%	3.8%	21.3%	2.2%
Group 2, Medium	22.0%	7.1%	11.1%	5.5%	23.9%	4.3%
Group 3, Small	5.4%	5.5%	6.7%	5.2%	7.8%	2.8%
Groups 1 & 2 (Combined)	4.4%	4.0%	5.7%	3.3%	16.0%	2.3%
<b>Total</b>	<b>4.1%</b>	<b>4.1%</b>	<b>6.1%</b>	<b>3.9%</b>	<b>7.1%</b>	<b>2.2%</b>

Maximum margins of error were not calculated for non-dichotomous items for which maximum variances were not possible to estimate.

### Sources of Error in the HHIS 2014

#### Population Specification and Specification Errors

As noted earlier, survey post-processing and eligibility screening for the small institutions resulted in revisions of population specification and specification error with respect to the institutional type. These errors were addressed by

- Additional adjustments to the sample strata weights based on the universe specification and recalculation of within strata unweighted response rates; and
- The development of analytical groups based on the self-specified institutional type.

Analytical groups were defined based on institutions' self-reported primary functional type and the size of the institution.

#### Measurement Error

There were several sources of measurement error in the HHIS 2014 data, largely due to survey instrument ambiguity and vague questions. For example:

- Instructions in the collections data section of the survey requested that respondents enter "999" if the number of units was unknown. A 999 could also have been a valid

numerical entry. It appears post survey data processing was able to address this source of error.

- Related to the above issue, no gateway questions were used to ascertain whether an institution held a particular type of collection item (see pp. 59-61 of the questionnaire in Appendix A). Therefore, blanks left by respondents were interpreted as zeroes or “not applicable” except for the 330 institutions that provided no collections data, which were coded as missing on these data elements.
- Two items in the list of collections, “Archival records/manuscripts” and “Maps and oversized items” included the ambiguous measurement unit instruction as follows: “record in linear / cubic feet.” This was remedied by developing separate variables for each measurement unit, which were then used to compute the “Total unbound sheets” at the three measurement levels: items, cubic feet and linear feet.
- Some questions included vague terminology not sufficiently defined for respondents that resulted in unreliable subjective judgments. In one instance, this resulted in omission of the resultant unreliable variables from the PUD, as described in Section VII. Users are advised to refer to the survey documentation and consider such limitations in using these data.

### Technical or Data Handling Errors

There were also several sources of technical or data handling errors associated with the HHIS 2014 data, such as:

- Application of Office of Management and Budget (OMB) approved imputation methods to missing data associated with physical and digital collections items as shown on pp. 59-61 of the questionnaire in Appendix A combined with other data collection issues associated with these items were problematic. Therefore, imputation for missing data for physical and digital collections items on pp. 59-61 of the questionnaire was not performed.
- Imputations for other missing data such as for types of personnel assigned responsibility for care of collections (see p. 50 of the questionnaire in Appendix A) used within record data cross-referencing.

### Response Bias

The overall survey response rate – given the adjustment for the universe after deduplication and eligibility assessment of the frame – was 20%, far below OMB’s guidance<sup>2</sup> that surveys needed to attain at least an 80% response rate. Social scientists and statisticians have become increasingly concerned about the general decline in response rates on surveys, therefore a new indicators of representativeness provide an alternative way to determine the extent to which response bias might threaten validity. In additional, these new analytical methods provide ways to adjust sample weights for known sources of unrepresentativeness.<sup>3</sup> These new methods were deployed with the HHIS 2014 data.

---

<sup>2</sup> OMB Reference (2006). [https://www.ftc.gov/system/files/attachments/data-quality-act/standards\\_and\\_guidelines\\_for\\_statistical\\_surveys\\_-\\_omb\\_-\\_sept\\_2006.pdf](https://www.ftc.gov/system/files/attachments/data-quality-act/standards_and_guidelines_for_statistical_surveys_-_omb_-_sept_2006.pdf)

<sup>3</sup> National Research Council. (2013). *Nonresponse in Social Science Surveys: A Research Agenda*. Roger Tourangeau and Thomas J. Plewes, Editors. Panel on a Research Agenda for the Future of Social Science Data

The widely-used Representativeness indicator ( $R(\rho)$ ) based on Schouten, Cobben, and Bethlehem (2009) was computed as follows for each of the 15 original group by type strata:

$$R(\rho) = 1 - 2S(\rho),$$

where  $S(\rho)$  is given by:

$$S(\rho) = \sqrt{\frac{1}{N-1} \sum_{i=1}^K (N_i)(\rho_i - \bar{\rho})^2},$$

and  $\rho_i$  is the response rate of strata  $i$  within each of the  $K$  strata,<sup>4</sup> and the mean response rate is given by:

$$\bar{\rho} = \frac{1}{N} \sum_{i=1}^K N_i \rho_i$$

and

$$N = \sum_{i=1}^K N_i$$

Table 10 shows the weighted response rates ( $\rho$ -bar in the equations, above) and representativeness index ( $R(\rho)$ ) for each type of institution. As shown, the  $R(\rho)$  are relatively high, with all but that associated with Archives over 0.85. These results suggest that despite the low weighted response rates, the questionnaire respondents can be taken as highly representative of the population from which they were drawn. Furthermore, the very high  $R(\rho)$  associated with the Group 3 institutions stratified by region indicate that there is also not a threat of response bias associated with region.

**Table 10. Representativeness Indexes by Type of Institution**

Type of Institution	Population Size	Weighted Response Rate	Representativeness Index ( $R(\rho)$ ), including Region for Group 3	Representativeness Index ( $R(\rho)$ ), (Overall, all groups)
Archives	667	35.7%	N/A	0.682
Historical Societies	2,794	20.0%	0.877	0.866
Libraries	11,257	17.0%	0.916	0.846
Museums	15,204	22.3%	0.955	0.848
Scientific Collections	1,368	14.8%	0.928	0.942

Collection, Committee on Statistics. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

<sup>4</sup> Region was used to stratify the sample of institutions within the four strata for which samples of less than 100% were drawn, namely the 43,574 institutions classified as small museums, libraries, historical societies, and scientific collections. Representativeness indexes ( $R(\rho)$ ) were computed within these four strata and used to adjust weights, as appropriate, based on proportionate sampling within the six American Alliance for Museums (AAM) regions (i.e., New England, Mid-Atlantic, Southeast, Midwest, Mountain Plains, and West).

## Nonresponse Bias Survey Results

OMB requires that when the unit response rate is below 80%, Federal agencies are expected to include a plan for a nonresponse bias survey (NRBS). For the HHIS 2014 data collection, a NRBS was conducted between February 20, 2015 and March 9, 2015, whereby a shortened version of the main questionnaire was sent to all non-respondent large and medium institutions and to half of the small institutions that had not responded (plus 71 of these institutions<sup>5</sup>), including those that had originally refused to participate during the eligibility verification phase of the survey administration.

Eight questions from the original survey were included in the follow up survey (shown in Appendix E):

1. Governance of the institution
2. Collections held by the institution (11 separate items)
3. Mission of the institution includes preservation of its collections
4. The institution has a preservation plan for its collections
5. Status of the preservation plan
6. Assessment of the condition of collections
7. Existence and status of an emergency plan that includes the collections
8. Funds for preservation of the collections are specified in the annual budget.

Two issues associated with response bias are simultaneously controlled by analyzing data separately by type and size. As shown previously in Table 8, response rates varied greatly as did the relative sizes of the populations within the 15 original strata in the sampling plan. As shown in Table 11, only three of the original 15 strata yielded a sufficient number of NRBS respondents for meaningful analysis of differences between respondents and non-respondents, all in Group 3: Historical societies (n = 59); Libraries (n = 132); and Museums (n = 95) (which are highlighted with grey shading in Table 11). Together, these three strata account for 91% of the universe of U.S. collecting institutions. Pearson's chi-square tests of the difference between NRBS respondents and those that responded to the original HHI 2014 survey were performed. Table 12 summarizes the significant ( $\alpha=0.05$ ) and meaningful<sup>6</sup> results of these statistical tests; none of the other dependent variables were found to have statistically significant differences between respondents and non-respondents in these three largest strata included in the HHI 2014 study.

**Table 11. Nonresponse Bias Survey Respondents**

Group and Size	Archives	Historical Societies	Libraries	Museums	Scientific collections	Total
Group 1, Large	4	5	23	21	6	59
Group 2, Medium	4	1	45	28	4	82
Group 3, Small	24	59	152	95	41	371
<b>Total</b>	32	65	220	144	51	512

<sup>5</sup> No further explanation for these additional 71 cases was provided in the contractor (RMC) file documentation.

<sup>6</sup> In this context, a five percentage point gap between the responses of NRB and HHI respondents is taken as a "meaningful" difference.

Table 12 indicates there were only a handful of statistically significant differences between HHI 2014 respondents and non-respondents. There were two consistent biases across all three groups (i.e., small Historical societies, Libraries, and Museums). First, those who had not responded to the HHI 2014 were more likely than responders to have already completed a general condition assessment of their collections. This finding might be interpreted to suggest HHI 2014 overestimates preservation needs for these three types of small institution since a general condition assessment is one of the initial activities associated with preservation.

Second, small Historical societies, Libraries and Museums that held Art objects were more likely to respond than their peer institutions that did not hold Art objects. Users should consider these results when drawing conclusions based on analyses of data for small Historical societies, Libraries, and Museums. No adjustments to the weights were performed based on the results of the NRBS.

**Table 12. Summary of Nonresponse Bias Study Results**

<b>Dependent Variables</b>	<b>Group 3 Historical Societies</b>	<b>Group 3 Libraries</b>	<b>Group 3 Museums</b>
<b>Governance</b> (Question 1)	No statistically significant differences	No statistically significant differences	No statistically significant differences
<b>Collections Types</b> (Question 2, 11 types of collection items)	<b>HHI more likely</b> than NRB to report: <ul style="list-style-type: none"> <li>Born digital</li> <li><u>Art objects</u></li> </ul>	<b>HHI more likely</b> than NRB to report: <ul style="list-style-type: none"> <li>Moving image</li> <li>Photos</li> <li><u>Art objects</u></li> </ul> <b>NRB more likely</b> than HHI to report: <ul style="list-style-type: none"> <li>Recorded sound</li> </ul>	<b>HHI more likely</b> than NRB to report: <ul style="list-style-type: none"> <li>Moving image</li> <li>Born digital</li> <li>Digitized</li> <li><u>Art objects</u></li> <li>Historic and Ethno objects</li> </ul>
<b>Preservation mission</b> (Question 3)	No statistically significant differences	<b>HHI more likely</b> than NRB to report Mission included preservation	No statistically significant differences
<b>Preservation planning and budgeting</b> (Questions 4, 5, 8)	No statistically significant differences	No statistically significant differences	<b>HHI more likely</b> than NRB to report: <ul style="list-style-type: none"> <li>Formal preservation plan</li> <li>Preservation included in budget</li> </ul>
<b>General condition assessment</b> (Question 6)	<b>NRB more likely</b> than HHI to report a general condition assessment of at least a portion of collection	<b>NRB more likely</b> than HHI to report a general condition assessment of at least a portion of collection	<b>NRB more likely</b> than HHI to report a general condition assessment of at least a portion of collection
<b>Emergency / Disaster Preparedness</b> (Question 7)	<b>NRB more likely</b> than HHI to report having an emergency/disaster plan	No statistically significant differences	<b>HHI more likely</b> than NRB to report having an emergency/disaster plan

### Weight Development Process

Weights for the 15 main (group by type) strata were determined via the following steps:

1. Compute design weights equal to the inverse of the probability of a given unit being selected.
2. Adjust weights due to response propensity based on cell response rates (propensity is the inverse of response rate, see Table 8, based on the R(p) computational process described above).

3. Adjust weights due to eligibility rates<sup>7</sup> (see Table 7).
4. Adjust weights for the four sampled strata (Group 3 Historical societies, Libraries, Museums, and Scientific collections) for regional stratification of original sample.

Users who wish to generalize to the original population from which the respondents were drawn should use the variable named WGT. As a cautionary note, SPSS implementation of weights inflates standard errors. Therefore SPSS users are advised to use sensitivity analyses and compute standard errors outside of SPSS for statistical accuracy.

---

<sup>7</sup> The computation of the population adjustment applied the eligibility determination to the full original population. For Groups 1 and 2 and the Group3 Archives strata (i.e., 11 of the 15 strata) this may be sound. For the remaining four strata (i.e., Group 3 Historical societies, Librarie, Museum and Scientific collections) which were originally sampled and for which the response rates were low, the population adjustment based on self-reported information from the institutions may have more error.

## **VI. POST-COLLECTION DATA PROCESSING AND DATA CONSIDERATIONS**

---

This section describes post-collection processing by RMC. RMC checked data for completeness and consistency and took steps to compensate for incomplete data and anomalous responses. This section describes the steps and techniques that RMC utilized in the process, including data review, consistency checks, identification of anomalous data, and data imputation.

### **Data Review and Consistency Checks**

Once the data collection period ended, RMC conducted a series of data checks and imported the checked data file from the web-based mechanism to SPSS for subsequent cleaning. A total of 1,714 records had completed all of the questions in sections A-F4 and were included in the final data file. Of these 1,714 records, 330 did not answer the physical collections items F5a-j (see pp. 59-60 of the questionnaire in Appendix A) and 588 had not completed the digital collections section (items F5k on p. 61 of the questionnaire). Among these 588 institutions, it was possible to determine via responses to items D10 and D11 whether they held digital collections, thereby permitting imputation of “not applicable” on all section 5k items.

### **Coding of Categorical and Numeric Variables**

RMC reviewed frequency distributions for all survey items to ensure input values were acceptable and developed editing specifications to check for logical relationships across items. For example, the response to one item may result in skipped items which were checked and then assigned a value of “-1” to indicate a “valid skip.” An instance where numeric data was blank and could be interpreted to be zero was replaced with “0.” Unanswered questions were coded as missing data indicated by “-9” or multiple “-99” when appropriate.

### **Coding of Open-Ended Response Variables**

It should be noted that respondents unable to report the number of items in Section F5 at the subcategory level entered text into the broad category “other” field indicating that they were reporting total amounts. Data processing included reviewing the text and recoding it as “Total” for any broad category values under the “other” subcategory.

With the exception of the above, no data entered as text in the “other” questionnaire fields are included in the PUD due to disclosure risk and data quality.

### **Response Option of “Don’t Know”**

Institutions were given the response option of “Don’t Know” for the majority of questions to accommodate respondents who may not have had specific knowledge about certain preservation activities within their institution even though survey instructions encouraged respondents to gather information from different departments or staff as needed. This was more problematic for the medium (Group 2) or large (Group 1) institutions. Researchers should be aware that responses of “Don’t Know” may indicate an uninformed respondent rather than the institution as a whole. If a respondent answered “Don’t Know” to a main question, all related sub-questions were skipped and coded as “Valid skip”.

Respondents were asked questions about specific preservation activities and given a list of potential answers where one or all could apply. One option on the list was “Don’t Know.” If the “Don’t Know” option was chosen, all other options within the list were recoded as “Don’t Know.” This allows a researcher to analyze these questions with comprehensive responses. This process was applied to the following questions: D7, D8, D8a, D9, D9a.

### **Institutional Staff for Conservation/Preservation (D7, D7a, D7b, D7c)**

Respondents were asked to indicate what type of staff was responsible for preservation activities (D7) including paid full-time, part-time, volunteer full-time, part-time, staff from other departments, contractors, or “Don’t Know.” In addition, responses about the number of paid professional full-time equivalents (FTEs), paid support (definition included assistants, IT staff, contractors, etc.) FTEs, and unpaid FTEs were used to impute the categorical responses to D7. As previously noted, surveys submitted online included data checks, but internal consistency checks for paper surveys needed to be completed by hand. The following data edits were made based on these consistency checks:

- If D7 indicated having paid full-time or paid part-time or staff from other departments and the number of FTEs for paid professionals (D7a) or support (D7b) were unanswered, these cases were flagged to impute data for D7a and D7b.
- If D7 indicated having volunteer full-time or volunteer part-time staff and the number of FTEs for unpaid volunteers (D7c) was unanswered, these cases were flagged for imputing data for D7c.
- If D7 indicated having contractors and the number of FTEs for support staff (D7b) was unanswered, these cases were flagged to impute data for D7b.
- If D7a indicated non-zero professional FTEs or support FTEs (D7b) was non-zero and D7 did not indicate having paid full-time or paid part-time or staff from other departments, FTE data was examined and paid full-time/part-time staff from other departments were coded as “2, Yes, imputed” based on similar institutions.

### **Overall FTE Staff (G1a, G1b)**

RMC conducted a consistency check between the paid conservation FTEs reported (D7a + D7b) and the overall paid FTEs at the institution (G1a). Likewise, the conservation volunteer FTE data was compared to the overall volunteer FTE data (G1b). Cases where conservation FTEs were greater than overall FTEs were reviewed alongside overall budget data (E1) and conservation budgets (E4). Inconsistent data was flagged as anomalous.

### **Annual Operating Budget (E1) and Annual Operating Budget for Preservation (E4)**

Preliminary analyses indicated that 207 Group 1 and 2 responding institutions reported that the conservation/preservation budget (E4) was less than one percent of the total operating budget (E1). RMC contacted these institutions via email and telephone to verify the total operating budget and conservation/preservation budget data. Of the 114 replies, 63 (55%) confirmed their responses and 51 (45%) revised their conservation/preservation budget to include staffing costs. The conservation/preservation budget data for the 93 institutions not verifying their data was considered anomalous and data was imputed. In addition, a comparison was made between the reported conservation budget and the reported conservation paid professional and support staff. There were several cases where the reported budget was zero, however paid FTEs were indicated. These cases were flagged as anomalous for the conservation budget amount and data imputed. Based on these findings, users of the HHI 2014 data should be aware that the preservation budget data may be underestimated due to respondents’ not including staffing costs.

RMC conducted a comparison of the conservation budget (E4) reported with the overall institutional budget (E1) to ensure the conservation budget was not larger than the overall budget. Anomalous data was identified and flagged for imputation. Overall, the item response rate was 90% for the total operating budget and 70% for the conservation/preservation budget.

### **Data Imputation for Institutional FTE Staff (D11b, G1a, G1b), Annual Budgets (E1, E4), and Visitors (G2a, G2b, G2c)**

Respondents were asked to report overall institutional FTE staff (G1a, G1b). For cases where appropriate data values were missing or flagged as anomalous, the median FTE value defined by

budgetary size and self-identified type subgroup was imputed. The budgetary size category was defined using the operating budget or the overall paid FTE or the group designation. Another round of validity checks were performed once data was imputed. In the event the imputed data resulted in illogical data for associated variables such as overall FTE and conservation FTE, missing values were assigned to the imputed data variable (-999) and flag variable (9) whereas the reported value was kept intact.

The same process was conducted for imputing the operating budget and/or the conservation budget. Budgetary size and self-identified type was used to create subgroup data for imputing the median budget. If data imputation resulted in conservation budget data greater than operating budget, the imputed data value was assigned missing (-999) along with the flag variable (9).

Median data values were imputed for the FTE staff conducting digital preservation (D11b) if applicable and missing. Similarly, the median number of visitors by size and self-identified type subgroup were imputed if questions were not answered.

### **Digital Material Collections (F5k)**

Respondents were asked to report the total volume of Digital Material Collection subcategories along with the data measurement size (GB/TB/PB). For data consistency, all item amounts were converted to gigabytes (GB) however the data measurement size reported remains intact.

## VII. CONSTRUCTION OF THE PUD

IMLS constructed the PUD, which contains 1,714 cases and 260 variables and is available as SPSS or CSV files. Appendix F is the record layout for the datafiles (based on the SPSS file), Appendix G provides unweighted and weighted frequencies for the categorical variables, and Appendix H provides descriptive analysis for the continuous variables included in the datafile. Each case was assigned a unique identifier, which is the first variable (CASEID) in each row.

The original questionnaire had 95 questions that were used to code 494 variables, not counting additional paradata variables. This section describes the considerations associated with adding and removing variables. Additionally, further data editing necessary to protect respondent confidentiality, assure data quality, and facilitate accurate data analyses by users is also described in this section.

### Protecting Respondent Confidentiality

Respondents were promised confidentiality (see survey letter in Appendix A). As such, IMLS followed several strategies to assure respondents could not be identified in the PUD:

- Open-ended questionnaire items that could reveal institution-specific information are not included in the PUD (5 questionnaire items);
- Personal information about the responder is not included (6 questionnaire items);
- Recoded variables to avoid small cells (2 variables);
- Continuous variables were topcoded within institutional type (8 questionnaire items exclusive of collections data; 12 physical collections variables; and 11 digital collections variables);
- Suppressed information for cases for which the number of valid responses within institution type was smaller than 5;
- Location information was limited to AAM region of the parent institution.

### Variables Recoded to Avoid Small Cells

Table 13 summarizes the transformation associated with creation of four new variables associated with disclosure risk and analytical issues. The variable named **Group2R** is coded to facilitate analyses and avoid possible disclosure of small cells as was the variable **Gov3R**.

**Table 13. Variables Recoded to Avoid Small Cells (Disclosure Risk)**

Original Variable → New Variable	Original categories	New Categories
Group → Group2R (Reference: Appendix C)	1 = Large	1 = Large and medium
	2 = Medium	
	3 = Small	2 = Small
Govern → Gov3R (Reference: Question B3, p. 47)	1 = College, university	1 = College, university
	2 = Non-profit, non-government	2 = Non-profit, non-government
	3 = Federal	3 = Government
	4 = State	
	5 = Local (county or municipal)	
	6 = Tribal	

*Top-Coding for Annual Budgets (E1, E4), Institutional Staff (G1a, G1b), and Visitors (G2a, G2b, G2c)*  
 Additionally, since extreme values associated with the continuous variables could be used to identify specific institutions, top-coding was performed in accordance with practices at other Federal

agencies.<sup>8</sup> Critical values were established for each variable and for each institution type so that these extreme values could be topcoded. The median value for each variable (controlling for institution type) for the cases flagged as higher than the critical value was used as the topcode value.

Table 14 provides an overview of the variables that were topcoded, including the critical values and medians that we assigned for each of institutional type. Details about topcoding physical and digital collections variables are provided in the section, below, titled “Data Quality: Item Response Rates”.

**Table 14. Topcoding Details, Continuous Items in PUD (Except Collections Variables)**

Question	Institution Type	Critical Value	# Cases Topcoded	Median Assigned
<b>Dig_stf_fte_T: D11b - Number of professional, support and volunteer FTE digital curation/preservation staff at institution</b>				
	Archive	20	3	47.0
	Historical Society	10	3	40.0
	Library	45	3	80.0
	Museum	33	3	180.0
	Scientific Collection	4	3	13.1
<b>pdfstaff_T: G1_a - Total paid FTE staff</b>				
	Archive	100	3	650.0
	Historical Society	95	3	207.0
	Library	799	3	1,050.0
	Museum	750	3	2,080.0
	Scientific Collection	54	3	150.0
<b>unpdfstaff_T: G1_b - Total unpaid FTE staff</b>				
	Archive	16	3	113.0
	Historical Society	30	3	50.0
	Library	60	3	106.2
	Museum	581	3	900.0
	Scientific Collection	3	3	15.0
<b>ann_opbudg_T: E1 - What is the total annual operating budget of your institution</b>				
	Archive	\$85M	3	451,000,000.0
	Historical Society	\$12.5M	3	22,493,394.0
	Library	\$312.7M	3	598,400,000.0
	Museum	\$211.0M	3	252,600,000.0
	Scientific Collection	\$20.8M	3	100,000,000.0

<sup>8</sup> Census Bureau – topcodes associated with IPUMS Current Population Survey see: [https://cps.ipums.org/cps/topcodes\\_tables.shtml](https://cps.ipums.org/cps/topcodes_tables.shtml); Bureau of Labor Statistics, Consumer Expenditure Surveys see: [https://www.bls.gov/cex/pumd\\_disclosure.htm#Basic](https://www.bls.gov/cex/pumd_disclosure.htm#Basic); and U.S. Department of Housing and Urban Development, Office of Policy Development and Research (n.d.) “Documentation Describing the Topcoding and Other Confidentiality Measures that Occur on the American Housing Survey (AHS) Public Use Files.

Question	Institution Type	Critical Value	# Cases Topcoded	Median Assigned
<b>pres_opbudg_T: E4 - What is the annual budget for preservation activities at your institution?</b>				
	Archive	\$1.5M	3	\$3,500,000.0
	Historical Society	\$225,181	3	\$426,694.0
	Library	\$3.21M	3	\$10,158,264.0
	Museum	\$3.12M	3	\$3,700,000.0
	Scientific Collection	\$220,000	3	\$590,858.0
<b>onsite_visit_T: G2_a - Number of onsite visitors</b>				
	Archive	376,740	3	\$842,742.0
	Historical Society	102,000	3	\$105,347.0
	Library	6.67M	3	\$10,966,290.0
	Museum	5.85M	3	\$6,200,000.0
	Scientific Collection	26,110	3	\$30,000.0
<b>offsite_visit_T: G2_b - Number of offsite visitors</b>				
	Archive	99,420	3	900,600.0
	Historical Society	80,270	3	150,000.0
	Library	297,480	3	422,866.0
	Museum	915,680	4	1,500,000.0
	Scientific Collection	2,140	3	20,000.0
<b>online_visit_T: G2_c - Number of online visitors</b>				
	Archive	57.33M	3	87,436,250.0
	Historical Society	1.15M	3	1,500,000.0
	Library	51.34M	3	206,174,520.0
	Museum	31.67M	3	500,000,000.0
	Scientific Collection	250,690	3	280,000.0

*Location Information Limited to AAM Region*

Including state as a separate location variable posed a disclosure risk. Therefore, the parent institution state location was used to assign each case to one of six AAM regions, as shown in Table 15. Users should note that there remains important analytical problems associated with Region because of the small number of institutions within each region, type, and group. Comparisons of variables across regions, therefore, should be approached with caution.

**Table 15. AAM Region Definitions**

Code (region)	Region	States
1	New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2	Mid-Atlantic	Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania
3	Southeast	Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, West Virginia, Puerto Rico, U.S. Virgin Islands
4	Mid-west	Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, Wisconsin
5	Mountain Plains	Colorado, Kansas, Montana, Nebraska, New Mexico, North Dakota, Oklahoma, South Dakota, Texas, Wyoming
6	West	Alaska, American Samoa, Arizona, California, Guam, Hawaii, Idaho, Nevada, Northern Mariana Islands, Oregon, Utah, Washington

### **Data Quality: Item Response Rates**

Item response rates (IRR) for all variables included in the PUD are provided in Appendix I. OMB<sup>9</sup> guidelines indicate that variables should have an IRR, computed by excluding valid/legitimate skips, of 70% or higher. All of the variables in this PUD conform to this standard. As indicated, above, limited missing value imputation was performed; only imputations that followed sound logic in conformance with OMB guidelines. With the exception of the collections variables, discussed below, three variables were removed due to data quality (these were associated with FTE staffing reported for items D7a, D7b, and D7c).

Data quality concerns associated with the interpretability of the items due to vague question wording were prevented release of 12 variables associated with item D13b with the PUD (these were associated with sources of damage). Similar information is conveyed via respondents' answers to the previous question, D13a, with clearer terminology and interpretability about the possible sources of damage to the institution's collections.

Nearly half of the original variables (235 in all) were related to respondents' answers to questions F5a-k, where respondents reported about the number and condition of 48 separate physical collection types within ten broad types (including an "other" category in seven of these ten types); and 13 types of digital collections (including an "other" category).

Many respondents found the section on amount and conditions of collections challenging, especially due to time constraints or just not knowing the information. There were 330 institutions that provided no collections data whatsoever, these are all coded as "missing" on all collections variables. Data imputation methods originally planned and approved by OMB resulted in unreliable data when implemented with these variables given the reporting difficulties by respondents. Additionally, all variables that coded respondents' assessments of the collections condition were determined to suffer from significant data quality problems and, therefore, are not released with the PUD.

---

<sup>9</sup> OMB's guidelines for federal statistical surveys and reporting are available at the following address: [https://www.ftc.gov/system/files/attachments/data-quality-act/standards\\_and\\_guidelines\\_for\\_statistical\\_surveys\\_-\\_omb\\_-\\_sept\\_2006.pdf](https://www.ftc.gov/system/files/attachments/data-quality-act/standards_and_guidelines_for_statistical_surveys_-_omb_-_sept_2006.pdf)

The structure of question F5, shown on pp. 58-60 of the questionnaire in Appendix A, did not provide an unambiguous, clear “entry item” to permit respondents to report whether a specific collection type was held by their organization. Lacking such a clear indicator, institutions that reported collections data were assigned a zero for any of the 48 collections items on which there were blanks.

Only the most reliable data are released with the PUD, specifically, the totals associated with each of the main categories of collections items, including those totals that were found to have been included in the “Other” category. These are as follows:

- Tbooks\_U: Books and bound volumes (number of items)
- TUnbd\_items\_U: Unbound sheets (number of items)
- TUnbd\_CF\_U: Unbound sheets (cubic feet)
- TUnbd\_LF\_U: Unbound sheets (linear feet)
- TPhotos\_U: Photographic collections (number of items)
- TMvImg\_U: Moving image collections (number of items)
- TRcSd\_U: Recorded sound collections (number of items)
- TArt\_U: Art objects (number of items)
- THist\_U: Historic and ethnographic objects (number of items)
- TArch\_in\_U: Archaeological collections, individually cataloged (number of items)
- TArch\_bk\_U: Archaeological collections, bulk (cubic feet)
- TNatSc\_U: Natural science specimens (number of items).

**Table 16. Topcoding Details, Physical Collections Variables**

Collection Type	Institution Type	Critical Value	# Cases Topcoded	Median Assigned
<b>Tbooks_U: Books and bound volumes (number of items)</b>				
	Archive	70,207	3	83,308
	Historical Society	359,275	3	457,175
	Library	12,298,061	3	14,826,922
	Museum	337,000	3	642,400
	Scientific Collection	54,000	3	140,000
<b>TUnbd_items_U: Unbound sheets (number of items)</b>				
	Archive	32000	3	325,000
	Historical Society	400,000	3	1,000,000
	Library	2,509,500	3	10,000,000
	Museum	500,000	3	6,041,914
	Scientific Collection	20,000	3	60,952
<b>TUnbd_CF_U: Unbound sheets (cubic feet)</b>				
	Archive	201,000	3	600,000
	Historical Society	41,585	3	118,000
	Library	80,000	3	764,629
	Museum	11,110	3	16,741
	Scientific Collection	1,187	2	2,500*
<b>TUnbd_LF_U: Unbound sheets (linear feet)</b>				
	Archive	22,366	3	25,377
	Historical Society	15,108	3	25,572
	Library	256,590	3	435,840
	Museum	25,500	3	41,500
	Scientific Collection	2,520	3	6,740

Collection Type	Institution Type	Critical Value	# Cases Topcoded	Median Assigned
<b>TPhotos_U: Photographic collections (number of items)</b>				
	Archive	4,310,000	3	6,841,500
	Historical Society	1,811,600	3	8,022,000
	Library	10,337,199	3	31,000,000
	Museum	1,220,9448	3	3,037,900
	Scientific Collection	42,416	3	237,839
<b>TMvImg_U: Moving image collections (number of items)</b>				
	Archive	245,000	3	807,000
	Historical Society	4,260	3	33,154
	Library	179,524	3	506,193
	Museum	12,053	3	46,688
	Scientific Collection	220	3	260
<b>TRcSd_U: Recorded sound collections (number of items)</b>				
	Archive	66,010	3	145,308
	Historical Society	6,702	3	23,246
	Library	475,644	3	663,000
	Museum	9,635	3	22,518
	Scientific Collection	28	3	100
<b>TArt_U: Art objects (number of items)</b>				
	Archive	29,000	3	35,404
	Historical Society	12,950	3	24,932
	Library	60,705	3	251,060
	Museum	94,039	3	173,329
	Scientific Collection	2,250	2	4,250
<b>THist_U: Historic and ethnographic objects (number of items)</b>				
	Archive	18,800	3	22,313
	Historical Society	46,839	3	98,500
	Library	8,950	3	13,945
	Museum	844,329	4	2,951,967
	Scientific Collection	1,474	3	15,100
<b>TArch_in_U: Archaeological collections, individually cataloged (number of items)</b>				
	Archive	127	3	2,320
	Historical Society	80,000	3	900,000
	Library	2,800	3	12,000
	Museum	5,000,000	3	13,000,000
	Scientific Collection	420,000	3	3,000,000
<b>TArch_bk_U: Archaeological collections, bulk (cubic feet)</b>				
	Archive	SUP	NA	SUP
	Historical Society	1,200	3	1,800
	Library	192	3	265
	Museum	29,242	3	600,000
	Scientific Collection	9,760	3	15,150

Collection Type	Institution Type	Critical Value	# Cases Topcoded	Median Assigned
<b>TNatSc_U: Natural science specimens (number of items)</b>				
	Archive	40	2	50
	Historical Society	352	2	3,925
	Library	650	2	1,000
	Museum	33,442,000	3	40,140,000
	Scientific Collection	6,000,000	3	8,706,500

Digital collections items were reported in a table associated with question F5k shown on p. 61 of the questionnaire in Appendix A. These items suffered from the same data quality problems as those associated with the physical collections items described, above. Respondents' answers to items D10, "Does your institution preserve digital collections" and D11, "Does your institution digitize collections" were used to assign "Not applicable" codes (-1) to those institutions that said "No" on D10 and D11 and then did not report answers for F5k. As with the physical collections, blanks were coded as zero for digital collections because respondents were not asked an entry question about the presence of twelve different types of digital materials. The volume of 11 types of digital materials were retained as the following variables, with volume data converted to Gigabytes as described in Section V:

- img\_vol\_U: Images (e.g., JPEG, PNG, TIFF, RAW, GIF, BMP)
- txt\_vol\_U: Texts
- vid\_vol\_U: Video (e.g., WMV, MOV)
- aud\_vol\_U: Audio (e.g., WAV, MP3, WMA)
- web\_vol\_U: Web sites
- dat\_vol\_U: Data sets
- sft\_vol\_U: Software
- elc\_vol\_U: Electronic records
- exh\_vol\_U: Exhibit media
- geo\_vol\_U: Geospatial media (e.g., GIS data)
- crec\_vol\_U: Original catalog records.

As with other continuous variables, all 11 of the digital collections variables were topcoded to avoid disclosure risk. Details concerning the topcoding process are provided in Table 17. In addition to topcoding, limited data suppression was also implemented for three of the digital materials variables. Specifically, data associated with digital audio collections for scientific collections, for digital exhibit media for both historical societies and scientific collections, and geospatial data for historical collections were suppressed due to the presence of fewer than 5 nonzero valid responses for these variables within institution type.

Table 17. Topcoding Details, Digital Collections Variables

Collection Type	Institution Type	Critical Value	# Cases Topcoded	Median Assigned (Gb)
<b>img_vol_U: Images (e.g., JPEG, PNG, TIFF, RAW, GIF, BMP)</b>				
	Archive	153,600	3	102,400,000
	Historical Society	30,720,000	3	1,101,559,808
	Library	589,520	3	3,216,316,416
	Museum	232,723,456	3	401,408,000
	Scientific Collection	55,000	2	100,500
<b>txt_vol_U: Texts</b>				
	Archive	7,000	3	31,562
	Historical Society	1,000	3	100,000
	Library	49,262,592	3	54,097,920
	Museum	20,181	3	22,588
	Scientific Collection	500	3	1,024
<b>vid_vol_U: Video (e.g., WMV, MOV)</b>				
	Archive	65,804	3	133,120
	Historical Society	321	3	25,600
	Library	7,680,000	3	14,485,504
	Museum	819,200	3	4,096,000
	Scientific Collection	18	2	50*
<b>aud_vol_U: Audio (e.g., WAV, MP3, WMA)</b>				
	Archive	8,479	3	51,200
	Historical Society	300	3	20,480
	Library	20,480,000	3	34,895,872
	Museum	41,894	3	460,800
	Scientific Collection	SUP	NA	SUP
<b>web_vol_U: Web sites</b>				
	Archive	3,891	3	5,000
	Historical Society	100	3	600
	Library	22,217	3	599,040
	Museum	6,144	3	20,480
	Scientific Collection	300	2	500*
<b>dat_vol_U: Data sets</b>				
	Archive	200	3	6,144
	Historical Society	20	2	20*
	Library	2,048	3	15,360
	Museum	9,216	3	17,408
	Scientific Collection	500	2	1,024*
<b>sft_vol_U: Software</b>				
	Archive	16	2	150*
	Historical Society	24	2	24*
	Library	195	2	1,024*
	Museum	2,048	2	12,288*
	Scientific Collection	100	2	200*

Collection Type	Institution Type	Critical Value	# Cases Topcoded	Median Assigned (Gb)
<b>elc_vol_U: Electronic records</b>				
	Archive	15,360	3	1,024,000
	Historical Society	2,700	3	15,000
	Library	3,309,806	3	7,283,103,744
	Museum	10,759,949	3	398,638,080
	Scientific Collection	1,024	2	2,048
<b>exh_vol_U: Exhibit media</b>				
	Archive	80	3	120
	Historical Society	SUP	NA	SUP
	Library	200	3	2,048
	Museum	40,000	3	2,048,000
	Scientific Collection	SUP	NA	SUP
<b>geo_vol_U: Geospatial media (e.g., GIS data)</b>				
	Archive	2,048	2	5,317*
	Historical Society	SUP	NA	SUP
	Library	7,168,000	2	204,902,400*
	Museum	2,000	2	2,048*
	Scientific Collection	Not necessary, 3 cases had the same maximum		
<b>crec_vol_U: Original catalog records</b>				
	Archive	36,851	3	330,000
	Historical Society	64,500	3	163,759
	Library	896,000,000	3	3,840,000,000
	Museum	276,415	3	2,560,000
	Scientific Collection	15	2	65

\* When there were two cases above the critical value, the code associated with the second highest variable value was assigned as the topcode for the maximum value. This was generally necessitated by significant order of magnitude differences between the top 2-3 values and all others within the specific variable and institution type.

Due to too few cases with valid data, data associated with games and that which was included in the “other” categories are not released due to disclosure risk. However, as with the physical collections data, other information requested of respondents in the chart shown on p. 61 about the storage location and condition of digital collections was of poor quality and is not released with the PUD.

### Variables Included to Facilitate Analysis

Two variables beyond those already discussed were developed to facilitate user’s analyses of the data. One variable (staff\_none) was created to facilitate analysis based on respondents’ answers to question D7, “Which best describes your current institutional staff (G) for conservation/preservation? (Select all that apply).” Respondents selected among six different types of staffing (full time paid, full time unpaid, part time paid, part time unpaid, volunteer, consultants, or other department). The variable called **staff\_none** in the PUD provides a way for users to capture whether a particular respondent reported that none of any of these six types of staff were reported by a particular institution as performing conservation/preservation duties.

Additionally, survey question B1 asked each institution to identify its primary type from among 21 different categories of collecting institutions. These were recoded into five categories for the analytical variable **rself\_type** as shown in Table 18.

**Table 18. Assignment of Institutions to the Variable rself\_type**

Code (rself_type)	Institution Type	Question B1 Response Categories
1	Archives	Archives
2	Historical societies	Historical society (includes genealogical societies, historical associations)
3	Libraries	Public libraries, Academic libraries, Independent research libraries (includes state libraries & large federal libraries), Special libraries (includes law, hospital, religious, blind & handicapped libraries)
4	Museums	Historic house/site, History museum (includes living history), Art museum (includes art gallery, art center, or arts organization), Children's/youth museum, Natural History museum, Science/technology museum, General museum (collection represents 2 or more disciplines), Specialized museum (collection represents one discipline), Arboretum or botanical garden, Aquarium, Nature center, Planetarium or observatory, and Zoo
5	Scientific collections	Archaeological Repositories/Scientific Research Collections = Archaeological repository or research collection, Scientific research collection (includes agencies or university departments with scientific specimen/artifacts)