

CHAPTER 2

The Cost of Creating Digital Images and Metadata by Museums

Howard Besser and Robert Yamashita (eds.). *The Cost of Digital Image Distribution: The Social and Economic Implications of the Production, Distribution, and Usage of Image Data* (A report to the Andrew W. Mellon Foundation), Berkeley: UCB School of Information Management & Systems, 1998.

Overview

The Museum Educational Site Licensing Project (MESL) brought together seven repositories of cultural heritage information and seven universities to collaborate on the content, capture, distribution, and use of images and associated text from the participating museums. This chapter examines the amount of time that the museums spent on the project, and how that might influence potential future endeavors.

The repositories participating in the MESL Project (six museums and the Library of Congress) each sent images and descriptive metadata about each image to a central site at the University of Michigan, which then concatenated sets from all the museums and sent them to the MESL universities. Almost all of the descriptive metadata supplied by the museums came from existing collection management records, and the process of preparing these for export to Michigan generally involved extensive reformatting to adhere to the standards developed for the MESL Data Dictionary. The majority of images supplied by the museums also came from a pre-existing stock of either digitized images or transparencies that had to be reformatted to meet MESL standards.

The purpose of this chapter is to identify and examine the major cost centers involved in repositories producing digital images for distribution over a network. The cost centers are:

- *Content selection*—the process of determining which images to distribute
- *Image preparation*—creating digital images or adapting previously digitized images
- *Image transmission*—the process of sending image files to the central distribution center in Michigan
- *Data preparation*—extracting text or raw data from a museum’s collection information system and reformatting it to fit the MESL data dictionary and field structure
- *Data transmission*—the process of sending text files to Michigan
- *Other*— time spent by museum staff that is not otherwise accounted for in the cost centers, including administration, research, coordination, and supervisory time

About the Museums

In two successive years, the museums provided digitized images and descriptive texts, representing over 9,000 works (at least 500 from each museum each year), to the participating universities. In order to participate in the project, museums had to meet certain requirements. They were expected to have existing automated collection documentation and be able to conform to project standards for text and images.

Seven institutions produced images and data for the MESL project:

- Fowler Museum at UCLA
- George Eastman House
- Harvard University Art Museums
- Museum of Fine Arts, Houston
- National Gallery of Art
- National Museum of American Art
- Library of Congress

Museum	Number of Objects	Number of MESL Staff
Fowler	115,000	6
George Eastman House	400,000	2-3
Harvard	150,000	7
Houston	36,000	8 (with 6 others performing significant work)
National Gallery of Art	94,000	3
National Museum of American Art	38,000	8
Library of Congress	100,000,000	4

Table 1—Participating Museums

Towards the end of the project each repository (except for the Library of Congress, which participated as a “special observer,”) completed a survey, referred to as the MESL Museum Technical Report. The survey included questions about the implementation process (including decisions made, time spent on various steps of the process, and technical details) and solicited information about how the staff felt upon the completion of the project. This chapter focuses on the reports of the six museums, and will add information from the Library of Congress when we’ve been able to obtain it. For more detailed information about each museum, see **Appendix 2A—About the Museums**.

Hidden Costs

Some significant costs were not included in the responses to the MESL Technical Reports due to the constraints of the project. Most of these costs were “hidden” because they were part of the normal museum process and not performed by the MESL Project team. Many of these costs were incurred prior to the MESL Project, and therefore details for these were sketchy at best. Here we describe two critical hidden costs: cataloging and rights clearance.

Cataloging

Cataloging consists of entering data about an object into a collection management system. The cataloging process can be broken down into accession, registration, inventory, and updating. Basic cataloging tasks might include correspondence, measuring, and data entry into a collection management system.

A museum normally catalogs an artifact for its own ongoing internal purposes, and some of that catalog information may be critical for an image retrieval and distribution scheme. But it is unclear how much of the cataloging information is really necessary for a project like MESL. Collection management records can be quite involved and complex, and it is not unusual for them to have more than 100 fields. In comparison, the MESL data dictionary contained only 32 fields. It is difficult to assess the cost of creating a collection management record, as these records are built over time by a number of different contributors and records are always in different stages of completion. It would be impossible to assess the costs of just creating the portion of a collection management record that was needed by the MESL data dictionary.

In the MESL Technical Reports, the amount of time to catalog an object ranges from less than half an hour to over three hours. The reason for the variance in time is that some museums may have only included data entry in their estimate. Upon further inquiry, at least one museum claimed that basic cataloging, without additional research into an artist's biographical information or the historical context of the object, generally takes between four and eight hours. It can take another hour or so to perform basic background research about the artist, such as finding alternate names, verifying birth date, and recording sources of information.

Rights Clearance

The cost of rights clearance directly affects digital distribution, but the issue of digital distribution rights is still a confusing one. Museums are already grappling with rights clearance for digital images when trying to obtain rights to include images on their own Web sites and CD-ROMs. While rights clearance is critical in the distribution of networked images, the time constraints of the MESL project prompted most of the museums to work around any clearance issues. In general the MESL museums avoided distributing images from the 20th century where a living artist or estate may still hold rights. Even though most museums avoided rights clearance during the MESL Project, for our analysis we tried to estimate what rights clearing would have cost.

Rights clearance at the museums is often handled by a Rights and Reproductions Administrator and is usually done for specific projects, such as the production of catalogs, books, or CD-ROMs. In the MESL technical reports, estimates regarding the amount of time it takes to clear an image range from ninety minutes to over three hours per image, but upon further inquiry, the museums report even greater variation. Easy rights clearance processes can be as simple as one fax to and from the rights holder, while more involved processes can drag on for months. The National Museum of American Art claims that 35% of a full-time employee's time was spent on rights clearance. The Museum of Fine Arts, Houston reports that for a recent catalog for a single artist exhibit, 20-25% of a full-time employee's time was spent on clearing rights.

Rights clearance generally consists of researching the rights holder (when necessary), writing a letter requesting permission and describing how the image will be used, negotiating payment (when necessary), and documenting the results of the negotiation and the permission when granted. Some rights, such as displaying the image, or making postcards or promotional items, are obtained upon

acquisition of the image. Museums that want to display images in a book, catalog, CD-ROM or Web site must often obtain permissions at that time. The current digital environment, however, is spurring museums to request digital distribution rights upon acquisition of the image.

Fees charged by rights holders for the use of images depend on whether the project is educational or commercial, the scope of the project, and whether an artist, estate, or agency holds the rights to the image. For example, in a limited distribution of images, such as a catalog or CD-ROM, fees for permission can range from \$20 to \$240 per image.

Similarly, museums have to contend with permissions requests from the community at large, including scholars, arts publications, and other media. If they do indeed hold the rights to the image or artwork in question, museums generally charge a fee for permission to use reproductions of their images. Traditionally, the fee has varied from approximately \$5 to \$50 depending on whether the use is educational, nonprofit, or commercial. In the past decade, the number of permissions requests has increased, thereby requiring museums to spend more time responding to inquiries and granting permission. (Sorkow, 1997) By participating in MESL, museums sought to explore a means of aggregating time-consuming individual permissions requests from scholars and universities by using a blanket site licensing agreement for online image distribution.

Total Hours

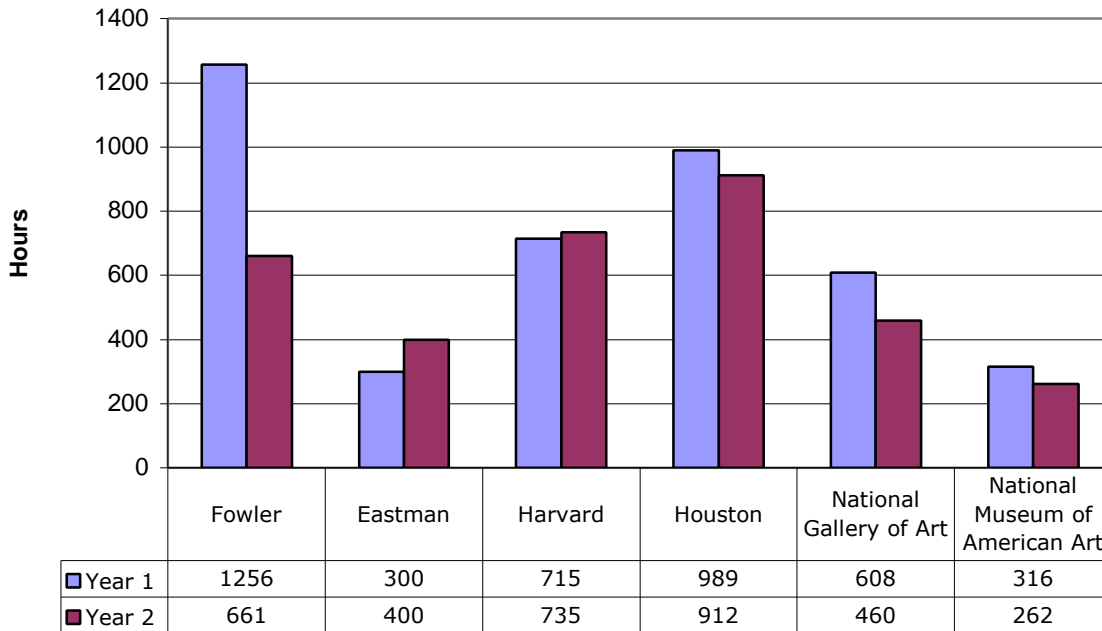


Chart 1: Total Hours

The total staffing hours reported by each museum for each year are shown in Chart 1, **Total Hours**. Total yearly hours ranged from a high of 1256 to a low of 262. In general, the number of hours contributed to the MESL project by each museum decreased in the second year. Museums cited the decrease in the learning curve when employees became familiar with the new technologies and MESL processes as reasons for the decrease in hours. Exceptions were due to the introduction of new technology or greater administrative time as reported in the Technical Report. The Fowler Museum spent the greatest number of hours in the first year due to numerous hours spent on image and data preparation (see the Image Preparation and Data Preparation sections below for more detail). We suspect that a third year would have shown slightly lower expenditures, and comments from the museums after the second year support this. Subsequent years would most likely level out around the Year 3 level if there were no significant technological changes.

Cost Centers

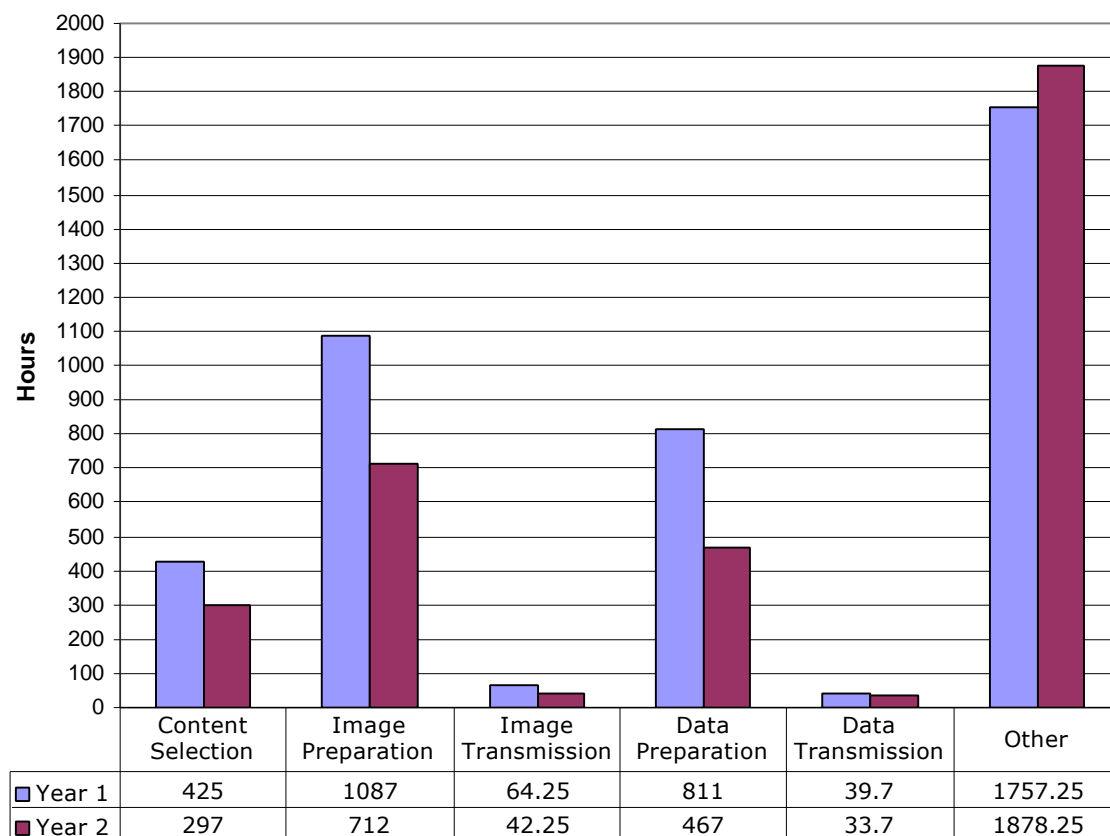


Chart 2: Total Hours for Each Cost Center

Chart 2, **Total Hours for Each Cost Center**, shows the total costs over all museums for each cost center for each year. The “other” cost center, representing administrative, supervisory, and research time, accounts for more than double the amount of time spent (over two years) in the next highest cost center, image preparation. Much of the time spent on “other” tasks were startup project costs such as attending meetings and planning strategies to complete the project. Although this is the one cost center in which hours increased in the second year, we estimate that the hours would decrease over time, as participants settled into standard procedures.

Overall, every cost center except for “other” decreased from Year 1 to Year 2. We believe that this trend would continue if there had been a Year 3, and that time spent on each cost center would remain relatively constant in subsequent years.

As shown in Chart 2, the image transmission and data transmission cost centers represent the smallest percentage of work spent by the museums. Technical professionals from the museums sent files to Michigan via FTP, recordable CD-ROM, DAT, or tape. The time involved ranged from under an hour to 25 hours. The variance in time reported is probably due to differences in exactly what the museums considered to be “transmission time.” Some museums may have included error correction (including changing file names) in their time estimates, while others reported only the precise time to transmit the files.

Content Selection

Images were chosen using various criteria, including:

- Availability of previously digitized images
- Objects/images without copyright issues
- Availability and quality of photography
- Quality and completeness of the text record
- Items on view
- Surveys to users
- Additional specific requests from universities

In the first year, the museums were under time pressure to get the images to the universities before the start of the fall semester so that faculty could use them for their courses.

In the second year, the museums tried to respond to feedback from instructors, who had indicated that they needed more involvement in the content selection process. Christie Stephenson and Thornton Staples from the University of Virginia created a Web-based request form to allow individuals to ask for specific images from the museums. Some museums displayed images on the Web, allowing instructors to view potential offerings (instead of just reading descriptions). (Notman, 1998)

For the most part, museums chose images that were specifically requested by instructors, that fit into their overall digitization plans, and that were already available. Museums avoided selecting images for which they were not the copyright holders.

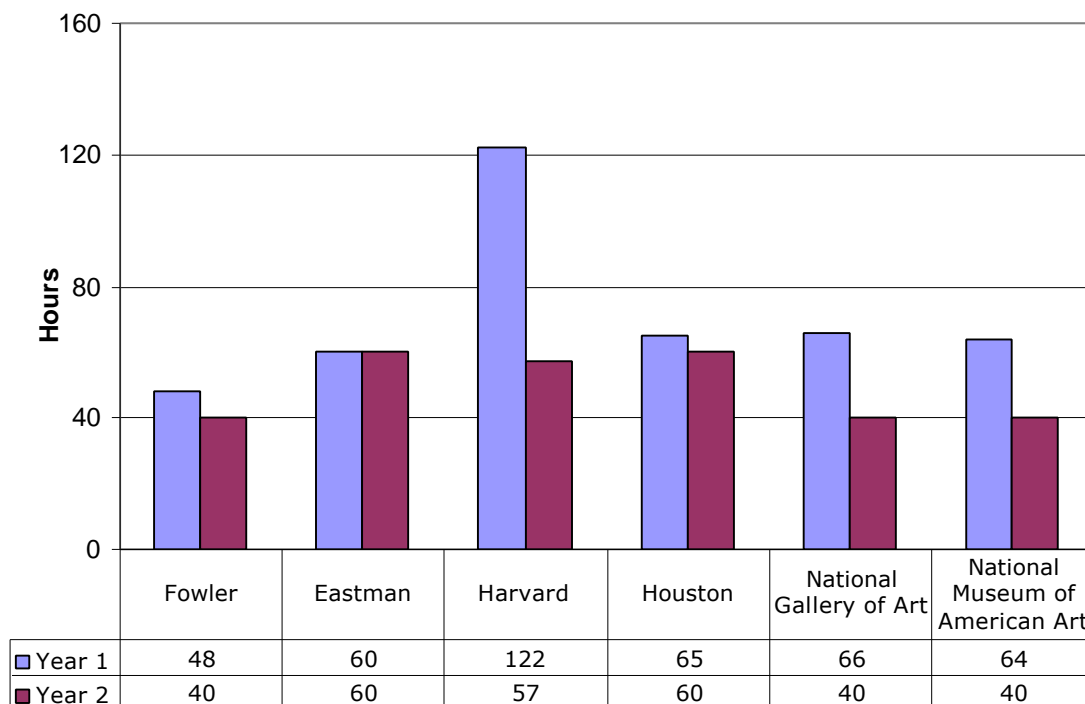


Chart 3: Content Selection

In general, the time involved in content selection decreased at each museum from the first year to the second (see Chart 3, **Content Selection**). Because the curators or project coordinators had already thought about which images to include in the first year, and due to use of the Web-based request form, the process of selecting images for the second year progressed faster than the first. Exceptions were due to internal administration issues. For more information on each museum’s content selection process, see **Appendix 2B—Content Selection**.

Image Preparation

Image preparation included creating digital images or adapting digital images from previous projects. The museums were requested to submit the highest quality images that they felt comfortable releasing. One concern that arose over the issue of image quality was the museums’ reluctance to release high-quality, easily reproducible images on the Internet. After some discussion, the museums realized that their concept of high resolution was different from the universities’ concept of high resolution. For the purpose of campus networks, high resolution meant 24-bit color, 1024x768 pixel images less than 3MB in size—resolution not high enough for quality reproductions. The museums felt comfortable with this level of resolution.

The process of digitizing images included scanning existing photos or taking new photos with a digital camera; adjusting the captured image for contrast, tonal balance, and scale; and saving the image in a file format such as JPEG, JFIF, TIFF, or GIF. Some museums outsourced images to a third-party vendor to be put on CD-ROM.

The breakdown in source media and digitizing devices used by the museums are shown in the following charts. Almost 50% of images were captured from large format film, while 35mm slides

or negatives, direct capture, and 35mm color copy negatives were the next most popular sources. The most popular digitizing devices were Photo-CD scanner and flatbed scanner.

Source Media for Digital Images	Percentage of Total Images
Direct capture	16%
35mm slides/negatives	20%
Medium-format film (4 x 5)	1%
Large-format film (8 x 10)	47%
Photographic prints	0% (1 image)
35mm color copy negatives	14%
Scanning from printed catalog	2%

Table 2—Sources of Digital Images

Digitizing Devices	Percentage of Total Devices
Video camera	1%
Digital camera	14%
Photo-CD scanner	34%
35mm film scanner	15%
Flatbed scanner	30%
Copy stand scanner	7%

Table 3—Devices Used to Capture Digital Images

Processes differed depending on whether the images were already digitized, came from catalogs, or were scanned from film negatives. Staff involved in image preparation were technical professionals, imaging specialists, work-study students, interns, and general photo services staff.

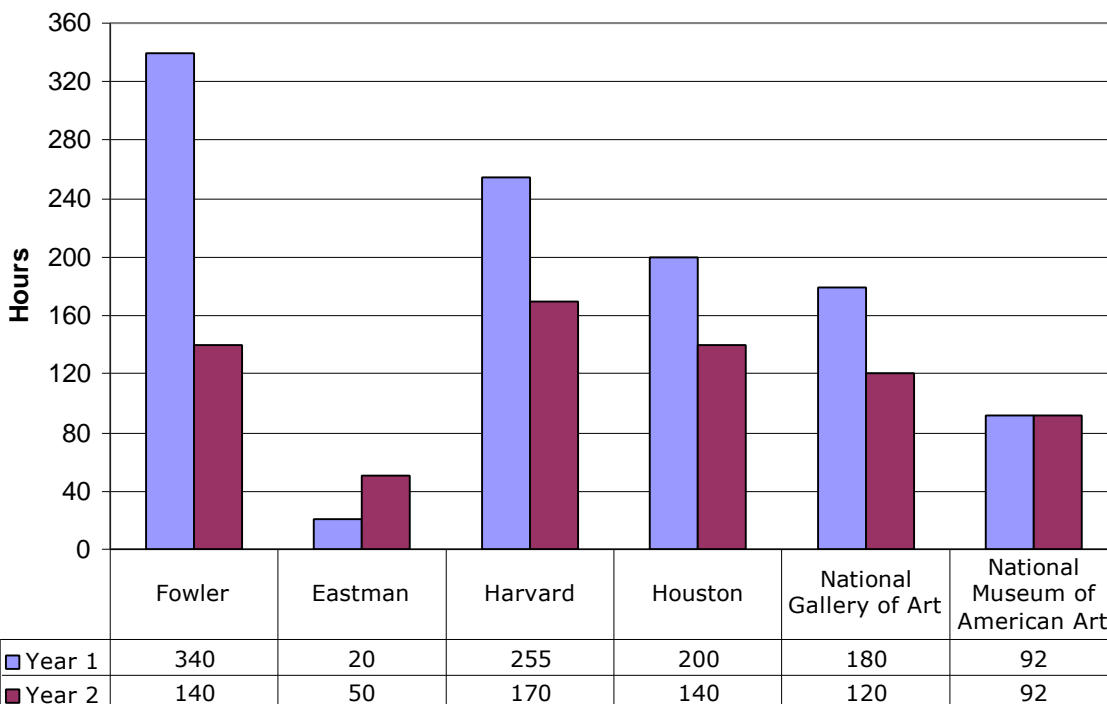


Chart 4: Image Preparation

As illustrated in Chart 4, **Image Preparation**, the average amount of time spent on image preparation in the first year was approximately 180 hours, while in the second year the average decreased to 120 hours. The decrease in the overall average is due to decreased learning curves and the museums’ increased familiarity with the MESL processes and technology. Eastman House had exceptionally low hours because most of the work was done for them at Kodak. The Fowler’s higher than average hours can be attributed to making high-resolution scans. For more detail about each museum’s image preparation process, see **Appendix 2C—Image Preparation**.

Data Preparation

Data preparation involved extracting text or raw data from a museum’s collection information system and reformatting it to fit the MESL data dictionary and field structure. All museums had technical professionals write macros to extract collections management information and program export routines and databases. In addition to technical staff, data preparation staff included curators, registrars, and interns.

Museums mapped their text descriptions to a data dictionary with 32 fields developed by the MESL project working group. Fields included accession number, creator name, creator place, material/medium, concepts/subject, description, and accompanying image. The list of fields can be found in **Appendix 2E—MESL Data Dictionary, Index of Field Names**. The complete MESL Data Dictionary, including explanations of each field, can be found in Appendix C of the MESL Final Report (Stephenson and McClung, 1998: 171-183).

The wide range in hours reported for data preparation illustrated in Chart 5, **Data Preparation**, (from a low of four to a high of 536) is not easily explained. The Fowler’s high number of hours may be due to additional staff members (more than the other museums), as well as a summer intern

working on data preparation in the first year. Except for their project coordinator, who was also the Director of Information Systems, the Fowler’s data preparation team was comprised of staff in non-technical positions. They performed a number of manual tasks in their data processing routine, including distributing data from one field into one or more applicable MESL fields and editing contents of all fields for consistency.

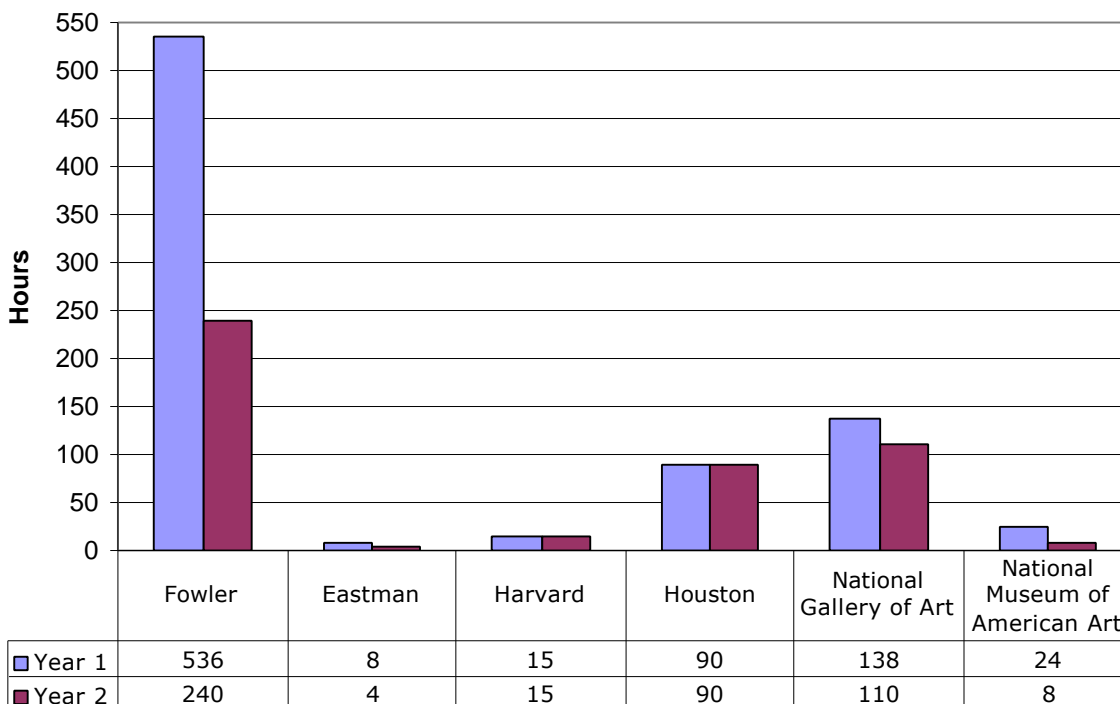


Chart 5: Data Preparation

Museums with significantly less time appear to be reporting just the amount of time to program an export routine. They either did not need to do manual cleaning of the files, or did not include that work in the report. From the data gathered, it looks as though the Fowler spent a considerable amount of time cleaning up their files, probably because their anthropological collection was the most difficult to map to the MESL Data Dictionary fields. For more details about the museums’ data preparation processes, please see **Appendix 2D—Data Preparation**.

Summary

This chapter has described the methods, personnel, tasks, and time necessary to produce digital images and related collection information for networked distribution. In addition to helping to summarize and determine the costs of the MESL project, this report should provide information for cultural heritage repositories or institutions that are considering digitization projects.

The greatest amount of time spent (except for administrative tasks) was for image preparation, which is the cost center most likely to see the greatest decrease in cost had this project continued for another year. As museums became accustomed to the technology, and found digitization routines that worked, the cost would have diminished as productivity increased. Advances in imaging technology continue to lower costs for computer software and hardware, and provide more

options for creating and manipulating digital images. Specifically, in MESL almost half the images digitized came from 8 X 10 transparencies. If this format is representative of what museums are using, then finding the most effective and efficient ways to scan large format film could result in significant cost savings.

The next highest amount of time spent was for text preparation. This cost center is somewhat more complicated because of the effort to map data to MESL specifications. Perhaps future projects will provide a different means of structuring the data that is more user-friendly for museums and other cultural institutions. In particular, data preparation time was significantly higher for the Fowler, an anthropology museum. It is likely that data preparation will be more complex, and the cost or time required much higher than average, for any museum that does not consider itself an “art” museum due to incompatible data requirements.

Another potential difference for future projects is content selection. The time required to select images may decrease if the process is not driven by user demand or request. Content selection may progress faster if images or objects are chosen in an order most convenient to the institution.

While time required for image and data transmission was negligible compared to the time required for the other cost centers, these activities will also see a decrease in cost as people become more familiar and comfortable with the technology, and settle on methods that work for all participants.