

David R. Meincke. Towards an Evaluation of Mass Digitized Photograph Collections. A Master's Paper for the MSLS degree. May 3, 2010. 54 Pages. Advisor: Stephen J. Fletcher.

This paper is a study in large collections of digitized photographs. The aim of the study is to determine what elements make up a collection that is mass digitized, and to generate a set of criteria for evaluating large digital photograph collections as mass digitization projects.

The study begins with a history of photographic archives and digital libraries, follows with a discussion of mass digitization, and ends with an evaluation of four collections: the *Keystone-Mast Stereograph Collection* at the University of California–Riverside; the *Los Angeles Times Photographs Collection* at UCLA; the *California Historical Society Photograph Collection* at USC; and the *Hugh Morton Collection of Photographs and Films* at the University of North Carolina–Chapel Hill.

#### Headings:

Photographic archives

Photograph collections—California—Los Angeles.

Photograph collections—North Carolina.

Digital libraries.

Digitized images

Digitization of archival materials

TOWARDS AN EVALUATION OF MASS DIGITIZED PHOTOGRAPH  
COLLECTIONS

by  
David R Meincke

A Master's paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Library Science.

Chapel Hill, North Carolina

April 30, 2010

Approved by

---

Stephen Fletcher

## Table of Contents

Table of Contents .....	i
List of Figures .....	iii
List of Tables .....	iii
1. Introduction .....	1
1.1 Historical Overview .....	2
1.2 Digitization .....	4
2. Mass Digitization .....	7
2.1 Light Processing .....	8
2.2 Benefits of Mass Digitization .....	9
2.3 Working Definition of Mass Digitization .....	10
3. Methodology .....	12
3.1 Metadata Quality .....	14
3.2 Randomization .....	14
3.2 Mass Digitization Index .....	15
4. Collections: Overview and Analysis .....	17
4.1 UC-Riverside California Museum of Photography: <i>Keystone-Mast Stereograph Collection</i> .....	19
4.1.1 Overview .....	19
4.1.2 Analysis .....	19
4.1.3 Sample Record .....	21
4.1.4 Data .....	22
4.2 UCLA: <i>Los Angeles Times Photographs Collection</i> .....	23
4.2.1 Overview .....	23
4.2.2 Analysis .....	24
4.2.3 Sample Record .....	25
4.2.4 Data .....	26
5. ....	26
4.3 USC: <i>California Historical Society Collection</i> .....	27

4.3.1 Overview .....	27
4.3.2 Analysis.....	27
4.3.3 Sample Record .....	29
4.3.4 Data .....	30
4.4 UNC- The Hugh Morton Collection of Photographs and Films (Collection P081, North Carolina Collection Photographic Archives, Wilson Library) .....	31
4.4.1 Overview .....	31
4.4.2 Analysis.....	31
4.4.3 Sample Record .....	33
4.4.4 Data .....	34
5. Conclusions.....	35
References .....	37
Appendix A: Study Methods and results .....	42
Appendix B: Assigning Random Numbers to Images .....	47
Appendix C: Weighting the MDI .....	49

## List of Figures

Figure 1. Record from UCR-CMP Keystone Mast Collection. “Gibraltar from a passing ship. The Straits, Mediterranean., March 5, 1931 .....	21
Figure 2. UCLA Library Digital Collections <i>Los Angeles Times</i> Photograph Collection. Collection Image Viewer with thumbnail view metadata of “ Los Angeles police officer inspecting area of Ambassador Hotel kitchen where Robert F. Kennedy was shot, 1968.” .....	25
Figure 3. Record for “Exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906.” California Historical Society Collection. USC Libraries Special Collections. University of Southern California. ....	29
Figure 4. Image Viewer and Metadata for “Grandfather Mountain, road to summit.” Hugh Morton Collection of Photographs and Films. University of North Carolina at Chapel Hill, Wilson Library, North Carolina Collection Photographic Archives....	33

## List of Tables

Table 1. Digitization Analysis of <i>Keystone-Mast Stereograph Collection</i> at UCR/CMP	22
Table 2. Digitization Analysis of Los Angeles Times Collection at UCLA .....	26
Table 3. Digitization Analysis of California Historical Collection at USC.....	30
Table 4. Digitization Analysis of Hugh Morton Collection at UNC .....	34
Table 5. Study Results .....	42
Table 6. Metadata Analysis for Keystone-Mast Stereograph Collection .....	43
Table 7. Metadata Variables for Keystone-Mast Stereograph Collection .....	43
Table 8. Metadata Analysis for Los Angeles Times Photograph Collection.....	44
Table 9. Metadata Variables for Los Angeles Times Photograph Collection .....	44
Table 10. Metadata Analysis California Historical Society Collection, 1860-1960.....	45
Table 11. Metadata Variables for California Historical Society Collection, 1860-1960..	45
Table 12. Metadata Analysis for Hugh Morton Collection of Photographs and Films ....	46
Table 13. Metadata Variables for Hugh Morton Collection of Photographs and Films...	46
Table 14. Weighting of Mass Digitization Index (MDI) .....	49

## 1. Introduction

Shortly after photography was introduced in the early nineteenth century by Joseph Niépce, Louis Daguerre and others, there was great uncertainty as how to use it as a cultural, historical, and scientific aid. Was photography simply a curiosity, a game played by the idle and chemically inclined? Or did photographic images of people, places, and events represent an incontrovertible, documentary truth? <sup>1</sup> One response to these questions came from François Arago, a mathematician and chemist, in an 1839 statement to the French Assembly. In it, he proclaimed the potential applications of photography, which ranged from studying and preserving the record of Egyptian hieroglyphics to mapping the stars. He listed other methods of preservation and exploration photography can engender, and then concluded with the statement that the potential for discovery in this new technology was larger than anyone could anticipate. He predicted that the inventors of photography, as the inventors of any new instrument of observation, would inevitably discover that what they “had hoped for from it [would seem] tiny in comparison to the succession of subsequent discoveries which it contributes.” <sup>2</sup>

---

<sup>1</sup> Some of these arguments as they relate to the early history of photography can be found in Walter Benjamin’s “A Short History of Photography.” A historical and theoretical text, this work discusses, among other things, the arguments of the proponents and detractors of the burgeoning photographic medium.

<sup>2</sup> Arago, François, “Report,” (1839) in *Classic Essays in Photography*, ed. Allen Trachtenberg (New Haven: Leete’s Island Books, 1980), p. 23.

Today, members of the information science community are in a similar position to that of the French Assembly in the early nineteenth century: a great method of preservation and discovery exists and yet the method and its implications are relatively unexplored, un-quantified, and even controversial. This new process that I am referring to is the mass digitization of photographs—the relatively high-speed and high-volume method of digitizing (scanning and making accessible through digital means) photographs with a focus on access and provision rather than description.

As advances are being made in scanning, storage, and network technology, photograph collections are becoming digitized at an increasingly rapid pace. With change comes controversy, however, and there are both criticisms and plaudits for mass digitization. Supporters of mass digitization cite preservation, access, and ease of scanning as reasons to digitize large portions of collections. Detractors of digitization question the effectiveness of digitization as a means to these ends.

One way of evaluating mass digitized collections for their effectiveness is to see how thoroughly they meet their objectives as mass digitization projects. As mass digitized collections are a relatively new concept, few methods exist for evaluation. This study aims to illuminate the variables and processes that contribute to an evaluation of a mass digitized image collection, so that evaluations of large digital image collections can be conducted in the future. Additionally, the study also aims to develop a working definition for the term, “mass digitization.”

## 1.1 Historical Overview

The study of digitized photographic collections is a combination of two topics within the broader discipline of information and library science: photographic archives

and digital libraries. Photographic archives, generally speaking, aim to preserve, organize, and provide access to photographic materials. Digital libraries, on the other hand, are harder to define. DELOS: an Association for Digital Libraries defines digital libraries as:

A term [which] has been generally used to refer to systems that are heterogeneous in scope and provide diverse types of functionality. These systems include digital object and metadata repositories, reference-linking systems, archives, content administration systems (mainly developed by industry), and complex systems that integrate advanced digital library services (mainly developed in research environments).<sup>3</sup>

Not surprisingly, the development of scanning technology has dovetailed with the growth of digital libraries. The first scan of a photograph was made in 1957,<sup>4</sup> but it took decades for scanning, computing, and network technology to progress to a state where it was both worthwhile and affordable for archival institutions to create digital libraries of scanned photographs. The three developments in scanning and computing technology that were necessary were: a high-enough quality of scan for the digitized image to be a faithful representation of the analog source; the availability of affordable and voluminous storage media; and the development of the Internet and subsequent exponential increases in network speed. These developments began to occur in the 1990s, and libraries quickly took advantage of this technology to create digital image libraries.<sup>5</sup>

Libraries of all sizes and types—academic, public, and even national libraries—began developing digital photographic libraries individually, with little standardization or inter-institutional communication. These digital photographic libraries were often little more than “a set of scanned photos posted within a locally designed Web homepage and

---

<sup>3</sup> Candela, L. et al. “The DELOS Digital Library Reference Model-Foundations for Digital Libraries.” DELOS: a Network of Excellence on Digital Libraries (February 2008): p. 3.

<sup>4</sup> Russell Kirsch of the National Bureau of Standards (known today as the National Institute for Standards and Technology, NIST) created the first scanned image, using a machine termed a ‘rotating drum scanner.’

<sup>5</sup> ZICK, G. “Digital Collections: History and Perspectives.” *Journal of Library Administration* 49.7 (2009): p. 688



various navigation designs.”<sup>6</sup> The move to standardization slowly occurred as a result of professional cooperation, increased awareness of photograph scanning, and guidelines set forth by large digital library groups and agencies [The Joint Information and Systems Committee (JISC), DELOS, the Library of Congress, etc.].

The Google Books Library Project has done much to propagate the idea of large-scale digitization. This project began in 2004, and aims to digitize and provide online access (when legal and possible) to books in a large number of research libraries. Prior to the Google Books’ digitizing, however, large-scale digitization projects were occurring with considerably less attention. Some institutions were digitizing their photographic materials in large amounts at a rapid pace, relying in part on established digitization conventions but also trying new techniques. These projects, of which the *Keystone-Mast Stereograph Collection* at the California Museum of Photography is one, represent some of the early successes and failures of digitizing large collections. In studying these early collections, one can determine how to plan and implement future digitization endeavors. The idea of “mass digitization,” could radically change the way digital image collections are conceived, processed, accessed, and digitized. It is important, therefore, to be able to define and evaluate them.

## 1.2 Digitization

There are many ways to digitize images. In an archival setting, these methods vary depending on institutional resources, the needs of the users, and the size the collection. Despite these differences, two basic components must be in place for the digitization process to occur. First, the library (or archive) needs a way of digitizing the

---

<sup>6</sup> Ibid, p. 689

images; generally, a scanner is the preferred method. Second, the digitizer needs to create a way of organizing the files so that each digital object possesses a unique identifier. After these steps have been taken, the digitization process is complete. It is not, however, a digital library yet. Next, the digitizer must create metadata for each object, and provide means of distribution for the digital images.<sup>7</sup>

In addition to these skeletal guidelines, there are other components of the digitization process. The National Archives and Records Administration's (NARA) guidelines are regarded as being among the most influential set of rules for digitization of objects. The guidelines state that digitization is:

“Not just as the act of scanning an analog document into digital form, but as a series of activities that result in a digital surrogate being made available to end users for a sustained length of time. The activities include:

- Document identification and selection
- Document preparation (including preservation, access review and screening, locating, pulling, and re-filing)
- Basic descriptive and technical metadata collection sufficient to allow retrieval and management of the digital copies and provide basic contextual information for the user
- Digital conversion
- Quality control of digital copies and metadata
- Providing public access to the material via online delivery of reliable and authentic copies
- Providing online ordering for reproduction services at quality or quantities beyond the capacity of an end user
- Maintenance of digital assets”<sup>8</sup>

Clearly, digitization consists of more than simply scanning. Thus, in digitizing large image collections, one can expect a greater deal of complexity. There are not many

---

<sup>7</sup> Astle, P. J., and A. Muir. “Digitization and preservation in public libraries and archives.” *Journal of librarianship and information science* 34.2 (2002): p 67.

<sup>8</sup> National Historical Publications and Records Commission. 2007. “Strategies and Tools for Archives and Historical Publishing Projects.” Grant announcement. <http://www.archives.gov/nhprc/announcement/strategies.html>.

resources available for aspiring mass digitizers of images, however, and there are even fewer scholarly studies on the subject. This difficulty makes the task of planning and implementing a mass digitized collection daunting and potentially problematical.

## 2. Mass Digitization

Throughout the library and archival communities, institutions are implementing mass digitization initiatives in various ways. As mentioned above, one of the more prominent uses of mass digitization is the Google Books Library Project.<sup>9</sup> Known for their scope and its implications for digital copyright, these projects have set standards for both large-scale digitization technology and practice. Another prominent mass digitization project is the Internet Archive. This project is not limited to books and includes digitized audio, video, and images in its collection.<sup>10</sup>

NARA is also working on a large-scale digitization project. This project aims to digitize a wide variety of materials, including photographs. Though not explicitly a mass digitization project, NARA's digitization initiative does have many elements in common with other mass digitized projects, including providing increased user access, enhancing preservation of records, and increasing the effectiveness of currently available resources. NARA's strategic plan illustrates their commitment to these aspects of mass digitization.

The plan includes the following objectives:

- Provide increased online access to an increased number of our holdings.
- Enhance preservation of records by reducing wear and tear on the originals for reference and reproduction.
- Provide access to those materials that can no longer be accessed in their original format.

---

<sup>9</sup> Google Books Library Project – An enhanced card catalog of the world's books.  
<http://books.google.com/googlebooks/library.html>

<sup>10</sup> The Internet Archive.  
<http://www.archive.org/about/about.php>

- Enhance users' understanding of records authenticity and archival context (e.g. who created the records; why were they created; how were they used).
- Use resources effectively. For example, original records that have been digitized may be relocated to less expensive archival storage locations. Partnerships, where the partner provides resources for digitizing, would expand the scale of digitizing beyond what NARA itself can do.
- Improve our service to customers consistent with their evolving expectations, and with consideration of NARA's available resources and customers' willingness to pay for value-added or convenience services.<sup>11</sup>

Many photographic archives, which often contain very large collections, have also turned to digitization to make material more accessible. Photographic materials, particularly negatives, can be difficult to use. Despite the relative ease of digitization, the majority of materials in most photographic archives have not been digitized. A possible solution to reduce this gap that is gaining ground is to employ the concept of mass digitization. Karen Coyle, in her article, "Mass Digitization of Books," defines mass digitization as "more than just a large-scale project. It is the conversion of materials on an industrial scale. That is, conversion of whole libraries without making a selection of individual materials."<sup>12</sup> This definition of mass digitization as it pertains to books can also be applied to photographic collections.

## 2.1 Light Processing

The concept of mass digitization has gained traction with the assistance of another idea: "light processing." The notion of "light processing"—conceptualizing archival materials on a macro-level and processing them quickly, rather than focusing on item-level specificity—was first described in 2005 by Mark Greene and Dennis Meissner in an

---

<sup>11</sup> National Archives and Records Administration. 2007. Plan for Digitizing Archival Materials for Public Access, 2007-2016." Draft version of 10 September, available for public comment online through 9 November 2007 at: <http://www.archives.gov/comment/nara-digitizing-plan.pdf>. [Accessed 11 February 2010]

<sup>12</sup> Coyle, K. "Mass digitization of books." *The Journal of Academic Librarianship* 32.6 (2006): 641.

essay “More Product, Less Process: Revamping Traditional Archival Processing.” Their intent was to bypass the “traditional” archival practice of item-level processing in favor of processing archival collections as minimally as possible in order to reduce “backlogs [that] are hurting the archival profession in the eyes of . . . researchers and resource allocators.”<sup>13</sup> In doing so, Greene and Meissner aimed to find a “Golden Minimum,” which would maximize access while keeping processing labor to a minimum.<sup>14</sup>

Mass digitization not only accommodates and implements the ideas of light processing, but it furthers its main tenet, that of access, by providing digital access to large quantities of material. In the article “Shifting Gears: Gearing Up to Get into the Flow,” the authors Erway and Schaffner describe how to process and digitize new materials in a fashion similar to that of light processing:

Stop obsessing about items. Everything that is digitized does not need to be painstakingly described. Archival control distinguishes organic collections from description of distinct books and museum objects. . . . While serious researchers value the description and organization that we bring to collections, after they’ve discovered a useful resource, they’re willing to contact us should they need more detail. . . . Vast quantities of digitized primary materials will trump a few superbly crafted special collections. Minimal description will not restrict use as much as limiting access to those who can show up in person.<sup>15</sup>

This model helps demonstrate how light-processing combined with mass digitization has the potential to be an excellent way of providing access to photographic collections that would otherwise remain unseen.

## 2.2 Benefits of Mass Digitization

Based on the preceding justifications and texts (“More Product Less Process,” the NARA Strategic Plan, “Shifting Gears”), the advantages of mass digitization over

---

<sup>13</sup> Greene, Mark, and Dennis Meissner. “More Product, Less Process: Revamping Traditional Archival Processing.” *American Archivist* 68.2 (2005): p. 2.

<sup>14</sup> Greene, Mark, and Dennis Meissner. “More Product, Less Process: Revamping Traditional Archival Processing.” *American Archivist* 68.2 (2005): p. 20.

<sup>15</sup> Erway, R., and J. Schaffner. “Shifting gears: gearing up to get into the flow.” Dublin, OH: OCLC Programs and Research. Retrieved November 5, 2007 (2007): p.6

limited, rich-metadata digitization are as follows: speed; preservation of materials; broader and more quality access; the provision of access to materials that cannot be accessed in original formats; an enhancement of users' understanding of contexts; and an effective use of resources.

Mass digitization is not for all collections, however. Some collections are more amenable to digitization than others are. Candidate collections for mass digitization are characterized as having noteworthy, rare, hard-to-obtain, or deteriorating materials. Other collections are not as amenable to mass digitization. These collections are characterized as having items that are generally accessible, common, or extremely difficult to scan.

The processor or project manager of a collection must carefully evaluate a collection before they decide to partake in a mass digitization project. Moreover, the institution must be willing to commit to the continual preservation of the digital resources. If the collection is not appropriate for mass digitization or the institution cannot commit to long-term preservation and access, then the institution should not undergo the significant effort of a mass digitization project.

### **2.3 Working Definition of Mass Digitization**

Defining mass digitization is one of the chief difficulties in studies on the subject. Dawn Schmitz writes that “the term mass digitization still has no universally accepted definition,” and that “‘mass’ digitization cannot be cleanly separated from ‘large scale’ digitization.”<sup>16</sup> However, for this study I will define mass digitization (as it applies to photographic collections) as a digital project that meets the following criteria:

---

<sup>16</sup> Schmitz, D. “The Seamless Cyberinfrastructure: The Challenges of Studying Users of Mass Digitization and Institutional Repositories.” Council on Library and Information Resources (2008): p. 4

- photographs are digitized rapidly;
- a high volume of photographs are digitized;
- photographs are digitized relatively indiscriminately (compared to smaller and more selectively digitized projects);
- volume and speed are favored over the quality of the metadata; and
- the digitization process is production-oriented, and there is a high degree of automation.



### 3. Methodology

Mass digitization is an evolving concept, and as a result it can be difficult to discern whether or not a collection was (or is in the process of becoming) mass digitized. To do this, I have attempted to identify key variables in existing collections that indicate whether a collection is mass digitized or not, and then create a formula using these variables to calculate a variable known as the “Mass Digitization Index” (see below). The concepts from which I will be drawing the variables are: the scale of the project; the quality of the images; the quality of the metadata; the rate of digitization; and the extent to which a digital collection refers to itself as a mass digitization project.

I have selected four digital photographic collections, noted for their sizes and differences from each other. Using freely available information to gather data such as dates, collection sizes, amount of digitized material, and image quality, I was able to develop evaluation metrics for the collections. I used the following variables to evaluate the collections:

**Digital Images (DI):** This refers to the number of digitized images available online through the digital collection.

**Total Images (TI):** This refers to the total number of analog images in the collection. This number is often approximate.

**Percent Digitized (PD):** This variable refers to the percent of the collection that is digitized.

**Display Resolution (DR):** This variable refers to the resolution of the display image. It is defined by the number of pixels on the long edge of the image.

**Archival Copy Resolution (AR):** This variable refers to the resolution of the archival copy of the image. It is defined by the number of pixels on the long edge of the image.

**Consciousness as a mass digitization project (CMD):** This variable registers the extent to which terms like “mass digitization,” appear in media (blogs, articles, conference papers) related to the digital collection. The value of this will be placed on a scale from zero to four. A value of zero means that the project has no public intention of being a mass digitization project. A value of four indicates that the collection fully identifies as a mass digitization project.

**Transparency (TY):** A subjective value, this refers to the extent that the collection makes its processes available to the online public. Blogs, news releases, and thorough collection descriptions and “about” pages contribute towards a high score in this category. This is an integer value between zero and four. A score of four means that there is an abundance of information on the collection. A score of zero means that no information is available.

**Year of Launch (YL):** This variable refers to the year in which the collection was made available online.

**Processing /Digitizing time (PT):** This variable refers to the time spent processing the materials, in years.

### 3.1 Metadata Quality

I also conducted a metadata evaluation of the collections based on individual records. From these collections, I randomly selected (using a random number generator) ten records and then evaluated those records. I used the following variables to evaluate the metadata:

**Metadata fullness (MF):** This variable will describe the extent to which the record is filled out. This value is an integer, consisting of a number between zero and ten. Zero indicates that the metadata record is blank, and ten indicates that the metadata record is as rich as reasonably possible.

**Ideal Metadata Fullness (IMF):** This value is an integer, consisting of a number between zero and ten. The highest MF value in the SRS of ten is used to provide this value.

**Metadata Average Quality (MAQ):** This value is the average of the ten metadata records' MF values in the SRS.

**Metadata Consistency (MC):** This value consists of the MAQ divided by the IMF. The closer it is to one, the more consistently attributed the metadata.

### 3.2 Randomization

Metadata quality is an inevitably subjective value. To forestall against surveyor bias, certain steps needed to be taken. The first step to combat bias is randomization. For each collection, ten numbers were generated using a random number generator (between one and the number of items in the collection). It was assumed that every item in the individual digital collection would have an associated unique number to which my randomly generated number would correspond. This did not necessarily prove to be true,

however, and each collection required a different method of randomization and selection. See Appendix B for the methods used to select random images from each collection.

The second step towards reducing bias is standardization. In my first effort towards standardization I created a conceptual “ideal record” for a photograph with all possible fields filled in. To create the ideal record, I took the Dublin Core Elements,<sup>17</sup> selected all the fields and elements applicable to photography, and filled in each of these fields. With this model in mind, I examined each photograph in my study and compared it to my ideal record. To allow for transparency and evaluation of my methods, I have included a table in Appendix A with the URLs and their respective metadata evaluation scores.

### 3.2 Mass Digitization Index

To synthesize my data and provide a variable that might serve to evaluate mass digitized collections in the future, I created a variable called the “Mass Digitization Index” (MDI). The MDI takes into account the size of the collection, the processing time, and the metadata level. The weighting of this index can be seen in Appendix B, but the justifications for the weighting of each variable are as follows.

First, minimal levels of metadata contribute to a high value of the MDI. As detailed metadata application—a process that includes research and manual transcribing on the part of the processor—is very time consuming, it takes away from the labor that could be used for digitizing. Second, both the size of the collection and the percent of the collection that is digitized bears heavily on the results of the MDI. I have also weighted

---

<sup>17</sup> Diane Hillmann. “Using Dublin Core.” *Dublin Core Metadata Initiative*. Web. 22 Feb. 2010. <http://dublincore.org/documents/usageguide/elements.shtml>

speed highly, as the aim of both mass digitization and light processing is to provide efficient and rapid access to materials. Finally, I have afforded a few points in the overall MDI to the CMD variable, which measures the extent to which a project incorporates the points I established in my definition of mass digitization.

The MDI is still an experimental variable, but one with much potential. With refinement, the MDI can be a useful tool to evaluate the effectiveness of large-scale digitization projects. Furthermore, future project managers and archivists will be able to better plan mass digitization projects by having a quantifiable method of examining the successes and failures of past projects.

#### 4. Collections: Overview and Analysis

I have analyzed four digital photographic collections, all of which feature large image collections with more than 5,000 digitized, web-accessible images:

1) The *Keystone-Mast Stereograph Collection* at the University of California at Riverside / California Museum of Photography. An example of a permanent museum collection comprising roughly 18,000 digitized images, this collection could be said to be mass digitized inasmuch as it a large-scale digitization project. Yet there are elements that partially stem from its location in a permanent museum setting (as opposed to an archive) that set it apart from the other collections in this study.

2) The *Los Angeles Times Photographs Collection* housed at the University of California at Los Angeles. This collection of just over 6,000 digitized images drawn from a physical collection of more than 1.5 million prints and 3.5 million negatives dating from 1918 to 1990 was released to the public in 2007.

3) The *California Historical Society Collection* at the University of Southern California. This collection of nearly 15,000 digitized items was made available digitally in the years 2003 and 2004. Like the *Los Angeles Times Photographs Collection* at UCLA, this collection was processed by and is housed by a large library that is part of a larger academic institution.

4) The *Hugh Morton Collection of Photographs and Films* at the University of North Carolina. This collection consists of about 500,000 items, and only a small portion

(roughly 4,500 at the time of the study) of the collection is available online through a web-accessible collection consisting of highlights from the collection.

## 4.1 UC-Riverside California Museum of Photography: *Keystone-Mast Stereograph Collection*

### 4.1.1 Overview

The California Museum of Photography (CMP) at the University of California–Riverside has an extensive collection of photographic collections. The collection that the Museum is arguably most famous for is the *Keystone-Mast Stereograph Collection*, the “world’s largest collection of original stereoscopic prints and negatives.”<sup>18</sup> The result of donations and acquisitions by the CMP, the collection contains nearly 350,000 items representing cultural and geographic history. Due to its historic nature, the fragility of the materials, and the completeness of the collection, this collection lends itself to mass digitization.

The idea for a large-scale digitization initiative for the *Keystone-Mast Stereograph Collection* was developed in the late 1990s. The initiative was part of a National Endowment for the Humanities project, known as the “Stereographs of the Americas” (SOA). This two-year project, which ran from 1998 to 2000, led to the creation of nearly 18,000 digitized images.<sup>19</sup> The CMP continued to digitize its collection after 2000. From 2000 to 2003, the CMP digitized just over 20,000 items. Currently, there are 42,778 items digitized, and the process is ongoing.<sup>20</sup>

### 4.1.2 Analysis

The “Stereographs of the Americas” project was not only effective, but it was pioneering in the field of mass digitization: in 1998, when the project began, the term

---

<sup>18</sup>Permanent Collections: *Keystone-Mast Stereograph Collection*. University of California, Riverside/California Museum of Photography, <http://www.cmp.ucr.edu/>.

<sup>19</sup> Permanent Collections: *Keystone-Mast Stereograph Collection*. University of California, Riverside/California Museum of Photography, <http://www.cmp.ucr.edu/>.

<sup>20</sup> Until 2013, according to the CMP web site.




mass digitization had not been conceived and Google Books, founded 2004, was still six years away.

The digitization for the *Keystone-Mast Stereograph Collection* occurred at a very fast pace. The first two waves of digitization—the first from 1998 to 2000 and the second from 2000 to 2003, resulted in 18,000 and 20,852 scanned images respectively, at a pace of 9,000 and 6,951 images digitized per year. The amount of metadata is relatively minimal; this points to a speedy digitization process, with an emphasis on access. The large number of digitized images and the relatively sparse metadata indicate that this is a project committed to mass digitization.

The inability to look at more than four images on a page, however, limits visual browsing capabilities that might greatly enhance the usability of the collection. This collection, with high MDI (a value of 85.7 for the first wave of digitization, and a value of 90.9 for the second wave of digitization) could benefit from an analysis of its user needs, and re-tailor its online interface to meet those needs.

### 4.1.3 Sample Record

COLLECTIONS PERMANENT  
KEYSTONE-MAST COLLECTION



1996.0009.KU95373.SS

George Lewis

Gibraltar from a passing ship. The Straits, Mediterranean., March 5, 1931  
 Keystone photo print  
 7.18 in. x 4.18 in.  
 Keystone-Mast Collection, UCR/California Museum of Photography, University of California, Riverside

Artifact Inscription:

[No inscription]

Location Information

Europe :: Gibraltar :: none

LC Subject Headings

Land :: Peninsulas  
 Gibraltar :: Vessels :: Ships

Figure 1. Record from UCR-CMP Keystone Mast Collection. “Gibraltar from a passing ship. The Straits, Mediterranean., March 5, 1931”<sup>21</sup>

<sup>21</sup> Record for “Gibraltar from a passing ship. The Straits, Mediterranean., March 5, 1931.” Permanent Collections: *Keystone-Mast Stereograph Collection*.. University of California, Riverside/California Museum of Photography,  
[http://www.cmp.ucr.edu/mainFrame/collections/guides/kmast/col\\_search\\_dis.lasso?-Search=Action&-Token.TokenDirect=front&-Table=online&-MaxRecords=1&-SkipRecords=103&-Database=collections&-Token.TokenURL=-MaxRecords%3d4%26-SkipRecords%3d100%26-token.TokenCategory%3dPhotography%26-Op%3deq%26keyword\\_combined%3dship%26-Op%3dcn%26event\\_ID%3d1996.0009%26-Op%3dbw%26object\\_category%3dPhotography&-token.TokenCategory=Photography&-KeyField=id&-Op=eq&keyword\\_combined=ship&-Op=cn&event\\_ID=1996.0009&-Op=bw&object\\_category=Photography](http://www.cmp.ucr.edu/mainFrame/collections/guides/kmast/col_search_dis.lasso?-Search=Action&-Token.TokenDirect=front&-Table=online&-MaxRecords=1&-SkipRecords=103&-Database=collections&-Token.TokenURL=-MaxRecords%3d4%26-SkipRecords%3d100%26-token.TokenCategory%3dPhotography%26-Op%3deq%26keyword_combined%3dship%26-Op%3dcn%26event_ID%3d1996.0009%26-Op%3dbw%26object_category%3dPhotography&-token.TokenCategory=Photography&-KeyField=id&-Op=eq&keyword_combined=ship&-Op=cn&event_ID=1996.0009&-Op=bw&object_category=Photography)

#### 4.1.4 Data

	<b>Keystone-Mast (first digitization wave, 1998-2000)</b>	<b>Keystone-Mast (second digitization wave, 2000-2003)</b>	<b>Keystone-Mast (second digitization wave, 2003-Present)</b>
Digital Images (DI)	18,000	20,852 (38,852 total)	3,426 (42,778 total)
Total Images (TI)	350,000	350,000	350,000
Percent Digitized (PD)	5.14%	5.96%	0.99%
Display Resolution (DR) (pixels on the long edge)	150 (prescribed), 500 (actual)	150 (prescribed), 500 (actual)	150 (prescribed), 500 (actual)
Archival Copy Resolution (AR) (pixels on the long edge)	3,000	3,000	3,000
Public Consciousness as a mass digitization project (CMD) (0 as unconscious, 4 as entirely conscious)	3	3	2
Transparency (TY) (0- 4)	2	2	1
Ideal Metadata Fullness (IMF) (0-10)	6	6	6
Metadata Consistency (MC) (0-10)	0.7	0.7	0.7
Metadata Average Quality (MAQ) (0-10)	4.2	4.2	4.2
Year of Launch (YL)	2000	2003	Ongoing (2010)
Processing /Digitizing time (PT) (years)	2	3	7
Speed (SP) (images/year)	9000	6951	489.4
Mass Digitization Index (MDI)	85.8	90.9	9.3

**Table 1. Digitization Analysis of Keystone-Mast Stereograph Collection at UCR/CMP**

## 4.2 UCLA: *Los Angeles Times Photographs Collection*

### 4.2.1 Overview

The *Los Angeles Times Photographs Collection* is one of the hallmark collections of the UCLA Digital Library. This collection consists of 6,227 digitized images taken from the *Los Angeles Times* Photographic Archives, which were donated to UCLA over the course of the 1980s and 1990s. The collection, according to the abstract, consists of negatives and prints documenting the cultural, political, and geographic history of Southern California and beyond.<sup>22</sup> The collection contains “more than 1.5 million prints and 3.5 million negatives from 1918 to 1990.”<sup>23</sup>

The digital collection, launched in 2007, contains 6,227 items, a seemingly large volume of photographs but actually only 0.12% of the entire physical collection. The project was digitized with a digitization staff of three, with support from fifteen catalogers, student assistants, and research assistants.<sup>24</sup>

The images are also viewable through a curated online collection called “Changing Times: Los Angeles in Photographs, 1920-1990.”<sup>25</sup> There are also high-resolution images from the collection available at the website for the Online Archive of California (OAC).<sup>26</sup>

---

<sup>22</sup> *Los Angeles Times Photographs Collection*. Collection Record. University of California, Los Angeles, <http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002np7z>

<sup>23</sup> Reynolds, Christopher. “Double exposure of history and art, in a shutter's click”. *Los Angeles Times*. 5 January 2007. <http://articles.latimes.com/2007/jan/05/entertainment/et-artphotos5>

<sup>24</sup> Report to the Digital Library Federation: Spring, 2004.” Digital library Federation. Council on Library and Information Resource. 1 February 2010. [http://www.diglib.org/pubs/news04\\_01/usc.htm](http://www.diglib.org/pubs/news04_01/usc.htm)

<sup>25</sup> “Changing Times: Los Angeles in Photographs, 1920-1990.” UCLA Digital Libraries. UCLA. 2 February 2010. <http://unitproj.library.ucla.edu/dlib/lat/index.cfm>

<sup>26</sup> “Finding Aid for the Los Angeles Times Photographic Archives, ca. 1918.” The Online Archive of California. California Digital Library. Web. 1 Mar. 2010. <http://www.oac.cdlib.org/findaid/ark:/13030/ft1q2n999m/>

### 4.2.2 Analysis

The *Los Angeles Times Photographs Collection* is particularly strong in its representation of the entertainment industry, and contains many images of some of the most recognizable public figures of the twentieth century. Despite its glitzy subject matter, however, the interface is relatively sparse. The collection lacks an introduction page, and as a result, usability suffers. This can be confusing to the user, particularly when the collection is also partially available through the “Changing Times” exhibit and through the finding aid available on the Online Archive of California (OAC), both of which have more explanation and relatively user-friendly interfaces.

The MDI value of 15.7 for the *Los Angeles Times Photographs Collection* is relatively low. The collection follows a relatively traditional model of metadata and browsing and provides a relatively low number of images. With 99.9 % of the collection not yet digitized, there is almost certainly additional material in the physical collection that is both interesting to the public and relevant to the scope of the digital collection. It would be a great boon to users of the collection if the digital library at UCLA initiated a mass digitization project using the *Los Angeles Times Photographs Collection*.

### 4.2.3 Sample Record

Digital Collections will be unavailable from 10 p.m to 11 p.m. on Monday, March 22 due to server maintenance.



**UCLA LIBRARY | Digital Collections**

Featured Collections | Browse Collections | Search and Find | Thematic Collections | Projects and Collaborations | Feedback

Digital Collections > Los Angeles Times Photographs Collection

**IMAGE CONTENT**



Click image to zoom

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License](#).

**Metadata**

<b>TITLE</b>	Los Angeles police officer inspecting area of Ambassador Hotel kitchen where Robert F. Kennedy was shot, 1968
<b>ALTERNATE IDENTIFIER</b>	uclalat_1429_b616_239495-4 [Local]
<b>COVERAGE</b>	<a href="#">Los Angeles (Calif.)</a> [Geographic]
<b>DATE</b>	1968-06-06 [Normalized]
	June 6, 1968 [Creation]
<b>FORMAT</b>	1 photograph [Extent]
	b&w negative [Medium]
	35 mm. [Dimensions]
<b>NAME</b>	<a href="#">Hackley, Gene</a> [Photographer]
	<a href="#">Kennedy, Robert F., 1925-1968--Assassination</a> [Subject]
<b>PUBLISHER</b>	<a href="#">Los Angeles Times</a> [Publisher Name]
<b>RELATION</b>	<a href="#">Los Angeles Times Photographic Archive, Department of Special Collections, Charles E. Young Research Library, UCLA, [Is Part Of]</a>
<b>SUBJECT</b>	<a href="#">Crime scenes--California--Los Angeles County</a>
<b>TYPE</b>	<a href="#">news photographs</a> [Genre]
	<a href="#">still image</a> [Type Of Resource]

Figure 2. UCLA Library Digital Collections *Los Angeles Times* Photograph Collection. Collection Image Viewer with thumbnail view metadata of “Los Angeles police officer inspecting area of Ambassador Hotel kitchen where Robert F. Kennedy was shot, 1968.”<sup>27</sup>

<sup>27</sup> Image Viewer and Metadata for “Los Angeles police officer inspecting area of Ambassador Hotel kitchen where Robert F. Kennedy was shot, 1968.” Library Digital Collections *Los Angeles Times Photographs Collection* University of California, Los Angeles.  
< <http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002vq4k>>

#### 4.2.4 Data

Digital Images (DI)	6,227
Total Images (TI)	5,000,000
Percent Digitized (PD)	0.12%
Display Resolution (DR) (pixels on the long edge)	750
Archival Copy Resolution (AR) (pixels on the long edge)	Unknown
Public Consciousness as a mass digitization project (CMD) (0 as unconscious, 4 as entirely conscious)	1
Transparency (TY) (0-4)	0
Ideal Metadata Fullness (IMF) (0-10 ;)	5
Metadata Consistency (MC) (0-10)	0.85
Metadata Average Quality (MAQ) (0-10)	5.1
Year of Launch (YL)	2007
Processing /Digitizing time (PT) (years)	2 (estimated)
Speed (SP) (images/year)	3113.5
Mass Digitization Index (MDI)	15.6

**Table 2. Digitization Analysis of Los Angeles Times Collection at UCLA**

### 4.3 USC: *California Historical Society Collection*

#### 4.3.1 Overview

The University of Southern California has an impressive digital library, and one of their hallmark collections is the *California Historical Society Collection*. This collection documents the history of Southern California from 1860 to 1960, with strengths in a variety of areas, including industry, politics, and urban development. The California Historical Society placed this collection on long-term deposit at USC in 1990<sup>28</sup>, and the collection was digitized around 2003.<sup>29</sup>

According to the collection's description, the *California Historical Society Collection* contains more than 23,000 photographs,<sup>30</sup> yet the digital collection contains 25,028 images. One possibility for this confusion is that since the writing of the description, either USC or the California Historical Society acquired more photographs. Aside from this minor discrepancy, the *California Historical Society Collection* is a well-produced digital collection with rich metadata and a very user-friendly interface.

#### 4.3.2 Analysis

*The California Historical Society Collection* received the highest MDI value out of the four collections in the study. This score is not without qualifications, however. First, it scored particularly high because of its high digitization rate (the entirety of the

---

<sup>28</sup> About – California Historical Society.” *The California Historical Society Collection*, 1860-1960. USC Digital Library. University of Southern California. 5 February 2010.  
<http://digitallibrary.usc.edu/search/controller/collection/chs-m15009.html>.

<sup>29</sup> This project first appeared as part of the USC digital initiative in 2003-2004, as demonstrated by its 2004 Mention in the “Report to the Digital Library Federation,” where it appears for the first time in the semi-regular series of reports (it does not in the 2002 report).

<sup>30</sup> About – California Historical Society.” *The California Historical Society Collection*, 1860-1960. USC Digital Library. University of Southern California. 5 February 2010.  
<http://digitallibrary.usc.edu/search/controller/collection/chs-m15009.html>.



collection has been digitized). This (possibly) inflated value (see above) lessens the impact the collection's unusually rich metadata might have on the MDI.

This thoroughness of metadata should contribute to a lower MDI value. This occurrence raises the suggestion that the MDI needs adjusting to become a viable evaluation tool. As the relative level of metadata is often indicative of the manual labor afforded to digitize the images, the metadata variable should be given more weight.

### 4.3.3 Sample Record

University of Southern California

USC LIBRARIES

USC DIGITAL LIBRARY

Home

Browse

My Selections


Search History

About

Help

Search

Advanced Search



CHS-49457

Exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906

[About this collection >](#)  
[View other items in this collection >](#)

Description

Photograph of an exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906. A brick road lies in the foreground with the large mansion behind it at center. The mansion is bordered on all sides by a sidewalk which is contained by a metal fence with thick, stone poles. A staircase leads from the sidewalk to the arched entrance which is damaged from the earthquake. The walls behind the entrance have crumbled to the ground.

**Titles:**  
Exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906

**Record ID:**  
chs-m23939

Names & Dates

**Created:**  
1906

**Creator:**  
[photographer] Pierce, C.C. (Charles C.), 1861-1946

**Publisher:**  
[host] University of Southern California Libraries

Subject

**Topic:**  
[News](#) - [Earthquakes](#) - [Fire](#) - [Crocker Mansion](#) - [unidentified place](#)

**Coverage:**  
1906

**Places:**  
[cities] San Francisco  
[countries] San Francisco  
[states/provinces] California  
[countries] USA

Relationships

**Part Of:**  
[Collection](#) Title Insurance and Trust / C.C. Pierce Photography Collection  
[1906-1906](#)  
[project] USC

Physical

**Type:**  
images

**Format:**  
negatives (photographic)  
3 photograph : photonegative, b&w ; 13 x 18 cm.  
photographs

Access

**Rights:**  
California Historical Society

**Repository:**  
USC Libraries Special Collections  
Doherty Memorial Library, Los Angeles, CA 90089-0189  
[access/06-49457](#)

**Ordering:**  
Send requests to address or e-mail given. Phone (213) 821-2366; fax (213) 740-2343.

Identifiers

**Record ID:**  
chs-m23939

**Source:**  
[accession no.] 49457  
[call no.] CHS-49457

[Submit a record request](#) for the USC Digital Library.  
[Contact us](#) if you have any questions or feedback.

©1996 - 2009 University of Southern California

Figure 3. Record for “Exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906.” California Historical Society Collection. USC Libraries Special Collections. University of Southern California.<sup>31</sup>

<sup>31</sup>Record for “Exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906.” California Historical Society Collection. USC Libraries Special Collections. University of Southern California.

<http://digitallibrary.usc.edu/search/controller/view/chs-m23939.html>

#### 4.3.4 Data

Digital Images (DI)	25,028
Total Images (TI)	23,000 *
Percent Digitized (PD)	100.00%
Display Resolution (DR) (pixels on the long edge)	1,024
Archival Copy Resolution (AR) (pixels on the long edge)	3,000 <sup>32</sup>
Public Consciousness as a mass digitization project (CMD) (0 as unconscious, 4 as entirely conscious)	0
Transparency (TY) (0-4)	1
Ideal Metadata Fullness (IMF) (0-10)	10
Metadata Consistency (MC) (0-10)	0.88
Metadata Average Quality (MAQ) (0-10)	8.8
Year of Launch (YL)	2004
Processing /Digitizing time (PT) (years, estimated)	2 (estimated)
Speed (SP) (images/year)	12,514
Mass Digitization Index (MDI)	476.2

\*This number may be incorrect

**Table 3. Digitization Analysis of California Historical Collection at USC**

<sup>32</sup> The specifications for the *California Historical Society Collection* are indicated in another source, the “Archival Resources for Visual Culture research: An Introductory Guide,” made by the participants of “Mining the Visual,” a symposium held in 2004.

## 4.4 UNC- The Hugh Morton Collection of Photographs and Films (Collection P081, North Carolina Collection Photographic Archives, Wilson Library)

### 4.4.1 Overview

The Digital Library at the University of North Carolina at Chapel Hill holds many digital photographic collections, including the *Hugh Morton Collection of Photographs and Films*. Of the four collections examined in this study, it is the only collection to identify itself as an exercise in mass digitization. The processors of the collection convey information about the collection through a blog, *A View to Hugh*.<sup>33</sup> In the blog, the processors write about the people, places, and events documented in the collection. They also provide details about technical, “behind the scenes” aspects of processing the collection. Though it is only partially processed, more than 4,500 selected images (at the time of this study) are fully digitized and online accessible.

### 4.4.2 Analysis

One way in which the *Hugh Morton Collection of Photographs and Films* differs from the other projects in this study is in how it utilizes the ideas espoused in Greene and Meissner’s “More Product, Less Process” through its use of scanning as a processing tool.<sup>34</sup> Another way in which the *Hugh Morton Collection of Photographs and Films* is unique among the collections in this study is in its public image and transparency of information, as well as its relatively high quality of metadata. Part of the reason the metadata quality is so high is due in part to volunteer/community involvement with the

---

<sup>33</sup> Hull, Elizabeth. “Mass digitization = mass confusion *A View to Hugh*. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web. 2 Mar. 2010.

<sup>34</sup> Fletcher, Stephen. “200,000 slides, part 3.” *A View to Hugh*. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web 25 April 2010.  
<http://www.lib.unc.edu/blogs/morton/index.php/2008/01/200000-slides-part-3/>

project. By engaging in activities such as “crowd-sourcing,”<sup>35</sup> and the solicitation of comments on the *View to Hugh* blog and the digital collection, this collection has made use of its interactions with the public to save resources and enhance its metadata.

That the *Morton* collection scored lowest on the MDI is surprising in light of its status as one of the more prominent mass digitization projects. Because of the presence of the digital collection’s rich metadata, the *Hugh Morton Collection* received a lower MDI value than it deserves.

---

<sup>35</sup> Hull, Elizabeth. “Crowdsourcing: Another Method.” *A View to Hugh*. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web. 2 Mar. 2010.<http://www.lib.unc.edu/blogs/morton/index.php/2008/07/crowdsourcing-ids%E2%80%94another-method/>

### 4.4.3 Sample Record

**HUGH MORTON COLLECTION**  
of PHOTOGRAPHS & FILMS

search | my favorites | all sections home

Grandfather Mountain, road to summit



**Title:** Grandfather Mountain, road to summit  
**Description:** Infrared photograph of road to summit of Grandfather Mountain, NC.  
**Filename:** P001\_NT054\_0052N3.LIF  
**Date:** 1950s  
**Search by Decade:** 1950-1959  
**Subject (tgm):** Mountain Roads; Mountains  
**Subject:** Infrared photographs; Automobiles  
**Geographic:** Grandfather Mountain (N.C.); Linville (N.C.); Avery County (N.C.)  
**Creator:** Morton, Hugh Macfarlane, 1901-2006.  
**Form General:** Photographs  
**Form Specific:** Black-and-white sheet film negative  
**Size of Original:** 4 x 5 in. (101.6 x 127 mm)  
**Collection in Repository:** Hugh Morton Photographs and Films (P001); Finding Aid: <http://www.lib.unc.edu/hcc/pca/p001/morton/morton.html>  
**Local Identifier:** P001  
**Digital Collection:** Hugh Morton Collection of Photographs and Films  
**Repository:** University of North Carolina at Chapel Hill, Wilson Library, North Carolina Collection Photographic Archives  
**Host:** University of North Carolina at Chapel Hill  
**Usage Rights:** Copyright North Carolina Collection, University of North Carolina at Chapel Hill Library. The images in this collection are made available for use in research, private study, and teaching. Prior permission from the North Carolina Collection is required for any commercial use.

UNC  
UNIVERSITY LIBRARY

[Digital Object Identifier](#) | [UNC Libraries Home](#) | [UNC Home](#) | [Research Help](#)

Figure 4. Image Viewer and Metadata for “Grandfather Mountain, road to summit.” Hugh Morton Collection of Photographs and Films. University of North Carolina at Chapel Hill, Wilson Library, North Carolina Collection Photographic Archives.<sup>36</sup>

<sup>36</sup> Image Viewer and Metadata for “Grandfather Mountain, Road to Summit.” Hugh Morton Collection of Photographs and Films. University of North Carolina at Chapel Hill, Wilson Library, North Carolina Collection Photographic Archives . University of North Carolina at Chapel Hill  
[http://dc.lib.unc.edu/cdm4/item\\_viewer.php?CISOROOT=/morton\\_highlights&CISOPTR=3965&CISOBOX=1&REC=15](http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&CISOPTR=3965&CISOBOX=1&REC=15)

#### 4.4.4 Data

Digital Images (DI)	4,500
Total Images (TI)	500,000
Percent Digitized (PD)	0.90%
Display Resolution (DR) (pixels on the long edge)	flexible
Archival Copy Resolution (AR) (pixels on the long edge)	4,000
Public Consciousness as a mass digitization project (CMD) (0 as unconscious, 4 as entirely conscious)	3
Transparency (TY) (0-4)	4
Ideal Metadata Fullness (IMF) (0-10)	9
Metadata Consistency (MC) (0-10)	0.844
Metadata Average Quality (MAQ) (0-10)	7.6
Year of Launch (YL)	Ongoing (2007–2010)
Processing /Digitizing time (PT) (years)	2.25
Speed (SP) (images/year)	2,000
Mass Digitization Index (MDI)	10.7

**Table 4. Digitization Analysis of Hugh Morton Collection at UNC**

## 5. Conclusions

The concept of mass digitization presents a paradox for academic study and archival practice. It is a nebulous idea that requires further study for it to be a reliable practice for photographic digital libraries. And yet future studies will only be possible as more projects employ mass digitization measures.

Indeed, though digital libraries are frequently discussed in professional and academic literature, both photographic collections and large-scale digitization initiatives are relatively neglected. It is my hope that this paper will not only provide a measure with which collections can be evaluated, but stoke further interest in the question of what elements make up a mass digitized photographic collection.

The idea that a variable can describe the extent to which a collection can be considered “mass digitized,” can be difficult to convey, but I feel that the creation of the Mass Digitization Index was a good first step. With future refinements, archivists and scholars of digitization will be able to use the MDI to able to evaluate large photographic collections against the definition of mass digitization. The MDI also provides a way in which archivists can measure the success of their own mass digitization projects.

With further refinement, the MDI can become a valuable tool for evaluating and planning collections. There are a few ways in which the MDI can be adjusted so that it can serve as a viable collection evaluation tool. First, more information is required, particularly in the variables of time, dates, and technical specifications. To obtain this



information, it is necessary to contact the curators of the collections. One variable that requires more information in order to be useful is the variable of processing time. This variable, employed in the calculation of the MDI, currently does not take into account the amount of hours expended into the collection. Instead, it paints a crude picture, measuring simply when the digitization started and when it stopped.<sup>37</sup> A more accurate variable to use would be “labor hours.”

Second, the variables need to be weighted differently. Currently, collections with unusually high values in certain variables (particularly “percent digitized”) can have disproportionately high MDI values. A refined MDI would provide a more nuanced weighting, better reflecting the way in which the collection meets the criteria set by the working definition of mass digitization. Finally, some variables are unnecessary, and convolute the data. The variables “consciousness as a mass digitization project,” and “transparency,” though interesting intellectual exercises, serve little purpose in providing evaluative data for a collection.

As the hardware for digitization is becoming increasingly inexpensive and the ideas behind it are becoming more popular, more collections are likely to be digitized; it will therefore benefit both the academic and professional communities to have a way to quantify their efforts.

---

<sup>37</sup> For example, if two institutions worked on separate digitization projects for two years and one institution had two technicians scanning full-time and an archivist encoding metadata twenty hours a week, and another institution had a student scanning and encoding metadata ten hours a week, their processing time would be the same

## References

- “About – California Historical Society.” *The California Historical Society Collection*, 1860-1960. USC Digital Library. University of Southern California. 5 February 2010. <http://digitalibrary.usc.edu/search/controller/collection/chs-m265.html>
- Arago, Francois, "Report," (1839) in *Classic Essays n Photography*, ed. Allen Trachtenberg (New Haven: Leete's Island Books, 1980).
- “Archival Resources for Visual Culture Research: An Introductory Guide,” Mining the Visual Symposium. Visual Culture Project, at USC. Los Angles, 2004.
- Armitage, L H, and P G B Enser. “Analysis of user need in image archives.” *Journal of Information Science* (1997): Print.
- Astle, P. J, and A. Muir. “Digitization and Preservation in Public Libraries and Archives.” *Journal of Librarianship and Information Science* 34.2 (2002): 67. Print.
- Baker, Mary, Kimberly Keeton, and Sean Martin. “Why Traditional Storage Systems Don’t Help Us Save Stuff Forever.” HP Laboratories Palo Alto, 1005. <http://www.hpl.hp.com/techreports/2005/HPL-2005-120.pdf>
- Benjamin, W. “A Short History of Photography.” *Screen* 13.1 (1972): 5–26. Print.
- Borgman, C. L. “What are digital libraries? Competing visions.” *Information Processing and Management* 35.3 (1999): 227–243. Print.
- Brown, Adrian. “Selecting Storage Media for Long-Term Preservation.” London: The National Archives (June 19, 2003). [http://www.nationalarchives.gov.uk/documents/selecting\\_storage\\_media.pdf](http://www.nationalarchives.gov.uk/documents/selecting_storage_media.pdf)
- Cappoli, D.,S. Oatey, and L. Vinales-Simpkins. “Hello Dali: Training the End-User of a Digital Image Archive.” *Information Outlook* 1.5 (1997): 27–30. Print.
- “Changing Times: Los Angeles in Photographs, 1920-1990.” UCLA Digital Libraries. UCLA. 2 February 2010.<http://unitproj.library.ucla.edu/dlib/lat/index.cfm>
- Coyle, K. “Mass digitization of books.” *The Journal of Academic Librarianship* 32.6 (2006): 641–645. Print.

- Candela, L. et al. "The DELOS Digital Library Reference Model—Foundations for Digital Libraries." DELOS: a Network of Excellence on Digital Libraries (February 2008) ISSN 8044 (1818): 2–912335. Print.
- Enser, Peter. "The Evolution of Visual Information Retrieval." *Journal of Information Science*. 34.4 (2008): 531-546. Print.
- Erway, R., and J. Schaffner. "Shifting Gears: Gearing Up to Get Into the Flow." Dublin, OH: OCLC Programs and Research. Retrieved November 5.2007 (2007): 2007–02. Print.
- "Finding Aid for the *Los Angeles Times Photographic Archives*, ca. 1918." The Online Archive of California. California Digital Library. Web. 1 Mar. 2010. <http://www.oac.cdlib.org/findaid/ark:/13030/ft1q2n999m/>
- "Finding Aid for the *Los Angeles Times Photographic Archives*, ca. 1918-." UCLA Digital Collections. University of California, Los Angeles. Web. 25 Mar. 2010.
- Fletcher, Stephen. "200,000 slides, part 3." *A View to Hugh*. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web 25 April 2010. <http://www.lib.unc.edu/blogs/morton/index.php/2008/01/200000-slides-part-3/>
- Fletcher, Stephen. "The Kodak HR Universal Film Scanner" *A View to Hugh*. 12 June, 2009 North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web. 3 Mar. 2010.
- Garber, Sharon R., and Mitch B. Grunes. "The Art of Search: a Study of Art Directors." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Monterey, California, United States: ACM, 1992. 157-163. ACM. Web. 27 Jan. 2010.
- Gernsheim, A., and A. Gernsheim. *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era*. McGraw-Hill, 1969. Print.
- "Google Books Library Project – An Enhanced Card Catalog of the World's Books." Google Books. Google. <http://books.google.com/googlebooks/library.html>. 2010 February 20.
- Greene, Mark, and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68.2 (2005): 208-263. Print.
- "Guide to the *Keystone-Mast Stereograph Collection*, 1870-1963." The Online Archive of California. California Digital Library. Web. 1 Mar. 2010. <http://www.oac.cdlib.org/findaid/ark:/13030/ft1q2n999m/>
- "Guidelines for Handling Image Metadata." Metadata Working group. 2008-09-24. <http://www.metadataworkinggroup.org/specs/>. 2009-12-28.

Hillmann, Dianne. "Using Dublin Core." Dublin Core Metadata Initiative. Web. 22 Feb. 2010. <http://dublincore.org/documents/usageguide/elements.shtml>

Hull, Elizabeth. "Crowdsourcing IDs: Another Method." A View to Hugh. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web. 2 Mar. 2010.  
<http://www.lib.unc.edu/blogs/morton/index.php/2008/07/crowdsourcing-ids%E2%80%94another-method/>

Hull, Elizabeth. "Mass digitization = Mass Confusion." A View to Hugh. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web. 2 Mar. 2010.  
<http://www.lib.unc.edu/blogs/morton/index.php/2008/02/mass-digitization-mass-confusion/>

Hull, Elizabeth. "Who Am I?—Folk Music Edition" A View to Hugh. North Carolina Collection Photographic Archives. University of North Carolina—Chapel Hill. Web. 23 Mar. 2010.  
<http://www.lib.unc.edu/blogs/morton/index.php/2009/05/who-am-i-folk-music-edition/>

Image Viewer for "Exterior view of Crocker Mansion on Nob Hill, showing earthquake damage, San Francisco, 1906." California Historical Society Collection. USC Libraries Special Collections. University of Southern California.  
<http://digitallibrary.usc.edu/assetserver/controller/view/search/CHS-49457>

Image Viewer and Metadata for "Grandfather Mountain, Road to Summit." *Hugh Morton Collection of Photographs and Films*. University of North Carolina at Chapel Hill, Wilson Library, North Carolina Collection Photographic Archives . University of North Carolina at Chapel Hill  
[http://dc.lib.unc.edu/cdm4/item\\_viewer.php?CISOROOT=/morton\\_highlights&CISOPTR=3965&CISOBX=1&REC=15](http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&CISOPTR=3965&CISOBX=1&REC=15)

Image Viewer and Metadata for "Los Angeles police officer inspecting area of Ambassador Hotel kitchen where Robert F. Kennedy was shot, 1968." Library Digital Collections *Los Angeles Times Photographs Collection*. University of California, Los Angeles.  
<http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002vq4k>

"Introduction - Definitions and Concepts," Digital Preservation Coalition.  
<http://www.dpconline.org/advice/introduction-definitions-and-concepts.html>

*Los Angeles Times* Photographs Collection Record. UCLA Digital Collections. University of California, Los Angeles.  
<http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002np7z>

"Museums and the Online Archive of California (MOAC). UCR/California Museum of Photography University of California, Riverside." Museum and Online Archives

of California (MOAC). Web. 25 Jan. 2010.  
<http://www.bampfa.berkeley.edu/moac/classic/ucrreport.html>

National Archives and Records Administration. 2007. Plan for Digitizing Archival Materials for Public Access, 2007-2016." Draft version of 10 September, available for public comment online through 9 November 2007 at:  
<http://www.archives.gov/comment/nara-digitizing-plan.pdf>. 11 February 2010

National Historical Publications and Records Commission. 2007. "Strategies and Tools for Archives and Historical Publishing Projects." Grant announcement.  
<http://www.archives.gov/nhprc/announcement/strategies.html>.

NIST Tech Beat, May 27, 2007  
[http://www.nist.gov/public\\_affairs/techbeat/tb2007\\_0524.htm](http://www.nist.gov/public_affairs/techbeat/tb2007_0524.htm)

Permanent Collections: *Keystone-Mast Stereograph Collection*. University of California, Riverside/California Museum of Photography, <http://www.cmp.ucr.edu/>.

Permanent Collections: *Keystone-Mast Stereograph Collection*. Search. University of California, Riverside/California Museum of Photography,  
[http://www.cmp.ucr.edu/mainFrame/collections/guides/kmast/col\\_kmast\\_search.l](http://www.cmp.ucr.edu/mainFrame/collections/guides/kmast/col_kmast_search.l)  
 asso

Peters, M., and B. Mergen. "'Doing the Rest': The Uses of Photographs in American Studies." *American Quarterly* (1977): 280–303. Print

Pitt, Sharon P., Christina B. Updike, and Miriam E. Guthrie. "Integrating Digital Images into the Art and Art History Curriculum.." *Educause Quarterly* 25.2 (2002): 38-44. Print.

Record for "Exterior view of Crocker Mansion on Nob Hill, Showing Earthquake Damage, San Francisco, 1906." *California Historical Society Collection*. USC Libraries Special Collections. University of Southern California.  
<http://digitallibrary.usc.edu/search/controller/view/chs-m23939.html>

Record for "Gibraltar from a passing ship. The Straits, Mediterranean., March 5, 1931." Permanent Collections: *Keystone-Mast Stereograph Collection*.. University of California, Riverside/California Museum of Photograph  
<http://www.cmp.ucr.edu/>

Reynolds, Christopher. "Double Exposure of History and Art, in a Shutter's Click". *Los Angeles Times*. 5 January 2007.  
<http://articles.latimes.com/2007/jan/05/entertainment/et-artphotos5>

Schmitz, D. "The Seamless Cyberinfrastructure: The Challenges of Studying Users of Mass Digitization and Institutional Repositories." *Council on Library and Information Resources* (2008): p. 4

Schwartz, J M. "Records of simple truth and precision': photography, archives and the illusion of control." *Archivaria* (2000). Print.

Sutherland, Juliet. "A Mass Digitization Primer." *Library Trends* 57.1 (2008): 17-23. Print.

"The Times' 128-year history". *Los Angeles Times*. 2010. 20 Jan 2010.  
<http://www.latimes.com/about/mediagroup/la-mediacenter-milestones,0,1482627,full.story>

University of Southern California. "Report to the Digital Library Federation: Spring, 2004." Digital library Federation. Council on Library and Information Resource. 1 February 2010.  
[http://www.diglib.org/pubs/news04\\_01/usc.htm](http://www.diglib.org/pubs/news04_01/usc.htm)

Waters, D. J. "What are Digital Libraries?" *CLIR (Council on Library and Information Resources) Issues*, No. 4. 1998. Accessed 2 February 2010.  
<http://www.clir.org/pubs/issues/issues04.html>.

Zick, G. "Digital Collections: History and Perspectives." *Journal of Library Administration* 49.7 (2009): 687–693. Print.

## Appendix A: Study Methods and results

	<b>Keystone-Mast (first digitization wave, 1998-2000)</b>	<b>Keystone-Mast (second digitization wave, 2000-2003) (38852 total)</b>	<b>Keystone-Mast (second digitization wave, 2003-Present)(42778 total)</b>	<b>Los Angeles Times</b>	<b>California Historical Society</b>	<b>Hugh Morton</b>
Digital Images (DI)	18,000	20,852	3426	6,227	25,028	4,500
Total Images (TI)	350,000	350,000	350,000	5,000,000	25,028	500,000
Percent Digitized (PD)	5.14%	5.96%	0.99%	0.12%	100.00%	0.90%
Display Resolution (DR) (pixels on the long edge)	150 (prescribed), 500 (actual)	150 (prescribed), 500 (actual)	150 (prescribed), 500 (actual)	750	1024	flexible
Archival Copy Resolution (AR) (pixels on the long edge)	3,000	3,000	3,000	?	3,000	4,000
Consciousness as a mass digitization project (CMD) (0 as unconscious, 4 as entirely conscious)	3	3	2	1	0	3
Transparency (TY) (0-4)	2	2	1	0	1	4
Ideal Metadata Fullness (IMF) (0-10)	6	6	6	5	10	9
Metadata Consistency (MC) (0-10)	0.7	0.7	0.7	0.85	0.88	0.844
Metadata Average Quality (MAQ) (0-10)	4.2	4.2	4.2	5.1	8.8	7.6
Year of Launch (YL)	2000	2003	Ongoing (2010)	2007	2004	Ongoing (2007-2010)
Processing /Digitizing time (PT) (years, estimated)	2	3	7	2	2	2.25
Speed (SP) (images/year)	9,000	6950.7	489.4	3113.5	12,514	2000
Mass Digitization Index (MDI)	85.8	90.9	9.4	15.7	476.3	10.7

**Table 5. Study Results**

<b>Number</b>	<b>Metadata Fullness (MF)</b>	<b>Item</b>
15345	4	1996.0009.KU15345
1021	6	1996.0009.A1021
25687	5	1996.0009.WX25687
15256	3	1996.0009.X152562
7222	4	1996.0009.KU7222
7170	3	1996.0009.KU17170
33375	6	1996.0009.33375.SS
2544	4	1996.0009.12544
9320	4	1996.0009.G9320
15647	3	1996.0009.KU15647

**Table 6. Metadata Analysis for Keystone-Mast Stereograph Collection**

<b>Ideal Metadata Fullness (IMF)</b>	6
<b>Metadata Average Quality (MAQ)</b>	4.2
<b>Metadata Consistency (MC)</b>	0.7

**Table 7. Metadata Variables for Keystone-Mast Stereograph Collection**



Random Number	Metadata Fullness	URL
5333	6	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002sn57">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002sn57</a>
2639	5	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002r03k">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002r03k</a>
6039	5	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002pq29">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002pq29</a>
1941	5	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002q0qf">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002q0qf</a>
3217	4	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002v6jj">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002v6jj</a>
611	4	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002pspk">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002pspk</a>
6161	6	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002q68m">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002q68m</a>
5089	6	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002pgnq">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002pgnq</a>
3406	5	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002t2vt">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002t2vt</a>
4149	5	<a href="http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002tr4m">http://digital2.library.ucla.edu/viewItem.do?ark=21198/zz0002tr4m</a>

**Table 8. Metadata Analysis for Los Angeles Times Photograph Collection**

<b>Ideal Metadata Fullness (IMF)</b>	6
<b>Metadata Average Quality (MAQ)</b>	5.1
<b>Metadata Consistency (MC)</b>	0.85

**Table 9. Metadata Variables for Los Angeles Times Photograph Collection**

<b>Random Number</b>	<b>Metadata Fullness</b>	<b>URL</b>
7324	9	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m7324.html">http://digitallibrary.usc.edu/search/controller/view/chs-m7324.html</a>
8243	10	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m8243.html">http://digitallibrary.usc.edu/search/controller/view/chs-m8243.html</a>
2965	9	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m2965.html">http://digitallibrary.usc.edu/search/controller/view/chs-m2965.html</a>
2555	8	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m2555.html">http://digitallibrary.usc.edu/search/controller/view/chs-m2555.html</a>
2238	8	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m2238.html">http://digitallibrary.usc.edu/search/controller/view/chs-m2238.html</a>
19841	8	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m19841.html">http://digitallibrary.usc.edu/search/controller/view/chs-m19841.html</a>
23269	8	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m23269.html">http://digitallibrary.usc.edu/search/controller/view/chs-m23269.html</a>
5976	9	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m5976.html">http://digitallibrary.usc.edu/search/controller/view/chs-m5976.html</a>
24673	9	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m24673.html">http://digitallibrary.usc.edu/search/controller/view/chs-m24673.html</a>
27	10	<a href="http://digitallibrary.usc.edu/search/controller/view/chs-m27.html">http://digitallibrary.usc.edu/search/controller/view/chs-m27.html</a>

**Table 10. Metadata Analysis California Historical Society Collection, 1860-1960**

<b>Ideal Metadata Fullness (IMF)</b>	10
<b>Metadata Average Quality (MAQ)</b>	8.8
<b>Metadata Consistency (MC)</b>	.88

**Table 11. Metadata Variables for California Historical Society Collection, 1860-1960**

Random Number	Title	Metadata Fullness
3000	7	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3000&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3000&amp;CISOBX=1&amp;REC=1</a>
1392	8	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=1392&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=1392&amp;CISOBX=1&amp;REC=1</a>
928	8	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=928&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=928&amp;CISOBX=1&amp;REC=1</a>
4407	7	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=4407&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=4407&amp;CISOBX=1&amp;REC=1</a>
490	8	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=490&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=490&amp;CISOBX=1&amp;REC=1</a>
3755	7	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3755&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3755&amp;CISOBX=1&amp;REC=1</a>
909	9	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=909&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=909&amp;CISOBX=1&amp;REC=1</a>
422	7	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=422&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=422&amp;CISOBX=1&amp;REC=1</a>
3970	8	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3970&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3970&amp;CISOBX=1&amp;REC=1</a>
1976	6	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=1976&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=1976&amp;CISOBX=1&amp;REC=1</a>
1836	8	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=1836&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=1836&amp;CISOBX=1&amp;REC=1</a>
3927	7	<a href="http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3927&amp;CISOBX=1&amp;REC=1">http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&amp;CISOPTR=3927&amp;CISOBX=1&amp;REC=1</a>

Table 12. Metadata Analysis for Hugh Morton Collection of Photographs and Films

<b>Ideal Metadata Fullness (IMF)</b>	9
<b>Metadata Average Quality (MAQ)</b>	7.6
<b>Metadata Consistency (MC)</b>	0.84

Table 13. Metadata Variables for Hugh Morton Collection of Photographs and Films

## Appendix B: Assigning Random Numbers to Images

In the *Hugh Morton Collection*, part of the UNC Digital Library, each item is assigned a unique number, ranging from 1 to the highest item in the collection. It was therefore easy to find items based on a random number sampling. I simply took the URL for the item viewer,

[http://dc.lib.unc.edu/cdm4/item\\_viewer.php?CISOROOT=/morton\\_highlights&CISOPTR=3000&CISOBX=1&REC=1](http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/morton_highlights&CISOPTR=3000&CISOBX=1&REC=1), and replaced the number following “CISOPTR,” with the pertinent number.

For the *Keystone-Mast Stereograph Collection* at the California Museum of Photography, the task was slightly different. As browsing and searching the collection is much different than the others, and the file numbering system was not reflective of digitized items but of all the items in the collection, a modified random sampling was conducted. I searched by number for the random number, and if the random number between 1 and 42778 was contained within the item number for an item in the collection (e.g., a search for number 15345 would bring up item 1996.0009.KU15345) then I would select it, and continue; in many cases the number was not found in the collection; in these cases, I would use the next randomly generated number.

For USC’s *California Historical Society Collection* the search was similar to the *Hugh Morton Collection*; I was able to take the base URL for a digitized photograph (<http://digitallibrary.usc.edu/search/controller/view/chs-m27.html>) and replace the final part “m##.html”) with the random number, between 1 and 25031.

For the *Los Angeles Times Photographs Collection* at UCLA the image identification number is not as transparent. To find the image that corresponded to the random number, between 1 and 6,227, the action had to be taken while browsing all images in the collection. The URL for the browsing is:

<http://digital2.library.ucla.edu/viewItem.do?ark=21198%2Fzz0002np7z&pager.offset=50&viewType=1&maxPageItems=50>

To obtain the image I needed, I changed the number in the string “pager.offset=50” to match my random number, and selected the first image.

Through this process, which at times felt like sleight-of-hand, and at others like fitting incorrect puzzle pieces into ill-placed, ham-fisted order, I was able to get a useable random sampling from these four collections.

## Appendix C: Weighting the MDI

$$\text{MDI} = (\text{DI}/400) + (\text{PD}/400) + (2\text{CMD}) + (\text{SP}/400) - (2\text{MAQ})$$

Digital Images (DI)	Divide by 400
Percent Digitized (PD)	400 x PD (PD of 5.14%, or .0514 would add 20.28 to the MDI)
Public Consciousness as a mass digitization project (CMD) (0 as unconscious, 4 as entirely conscious)	2x CMD
Metadata Average Quality (MAQ) (0-10)	2x Inversely Weighted (MAQ of 1 would add 20 to MDI; MAQ would add 10; MAQ of 10 would add 2)
Speed (SP) (images/year)	Divide by 400

**Table 14. Weighting of Mass Digitization Index (MDI)**