Reformatting Preservation Departments: The Effect of Digitization on Workload and Staff

Marie R. Kennedy

This study investigates whether digitization has affected the workload and staffing of preservation departments. Data from a survey of eighteen ARL libraries over five years were used to track the number of reformatting tasks completed and staffing trends in order to determine whether there is an evident effect from digitization. Analysis reveals that the number of items processed by preservation departments has increased by ten percent due to digital-reformatting tasks and without a corresponding increase in staffing. The shape of preservation departments is indeed shifting, and this trend should be followed closely over subsequent years

As the use of digital objects in libraries continues to rise, it is reasonable to evaluate the effect the increase has on the departments within the library that put those digital objects in place. This paper looks at the effect digitization has had on preservation departments in particular, as they complete many of the digital tasks of a library. The Association of Research Libraries (ARL) has identified preservation departments as significant producers of digital material and began tracking their digital output in 1998. This paper uses the data gathered by the ARL to track the progress of digitization within and across preservation departments and evaluates its effect on other functions of those departments. Actual digitization practices of a subset of preservation departments also are considered in re-lation to the data provided by the ARL preservation statistics.

The ARL is composed of 123 libraries in the United States and Canada. It is an institutional membership organization whose mission is to operate "as a forum for the exchange of ideas and as an agent for collective action."[1] This analysis will use the preservation statistics gathered by ARL libraries because they have been consistently gathered since 1998 and may provide evidence of shifting workloads and staffing over time.

Group 4 ARL Libraries

In 1991, Jan Merrill-Oldham, Carolyn Clark Morrow, and Mark Roosa outlined a theoretical program model for mature preservation programs that organizes the 123 ARL libraries into four groups, based on total number of volumes held.[2]

Group 4, characterized by holdings of more than five million volumes, has the greatest number of personnel and the biggest budget. In variation from the other three models described, Group 4 maintains a separate photoduplicating unit that performs reformatting tasks such as microfilming and photocopying. The other groups contract out their microfilming and photocopying; it is assumed that Group 4 has in-house facilities for this purpose.

In imagining an updated version of the Merrill-Oldham model, the photoduplicating unit can be viewed as an appropriate place to now also house digitization processes as this, too, is a reformatting task. This study uses this re-imagined Group 4, now including digitization, as the framework for this analysis.

Based on the ARL interactive statistics Web site (http://fisher.lib.virginia.edu/arl/index.html), the following libraries have been identified as having more than five million volumes (as of 2002, the latest statistical year available), thus placing them into Group 4: Arizona, British Columbia, California-Berkeley, California-Los Angeles, Chicago, Columbia, Cornell, Duke, Harvard, Illinois-Urbana, Indiana, Michigan, Minnesota, North Carolina, Ohio State University, Pennsylvania, Princeton, Stanford, Texas, Toronto, Washington, Wisconsin, and Yale. Of these twenty-three, Arizona, British Columbia, Minnesota, Pennsylvania, and California-Los Angeles do not have a full-time preservation administrator and are not considered in this analysis. The remaining eighteen Group 4 libraries considered here all have at least one full-time preservation administrator.

ARL Digitization Statistics

As mentioned earlier, preservation statistics have been gathered from ARL member libraries since 1988. Additional statistics related to digitization tasks have been gathered since 1998. This analysis examines all five years of preservation statistics that include digitization tasks

TABLE 1 Total Pieces Reformatted in Group 4 Libraries, 1998–2002	
Reformatting Task	**Items Completed**
Volumes photocopied	20,112
Volumes microfilmed	148,960
Volumes digitized	**51,254**
Sheets photocopied	545,069
Sheets microfilmed	6,469,401
Sheets digitized	**558,409**
Nonpaper items photocopied or microfilmed	32,212
Nonpaper items digitized	**198,296**

(1998–2002). Within these five years, several Group 4 libraries do not include data for digitization tasks. In 1998, for example, six of the eighteen Group 4 libraries did not report any volumes digitized, eight did not report any sheets digitized, and four did not report any nonpaper items digitized. It is not known by examining the data if this means that these libraries did not perform digitization tasks or if they simply chose not to report them.

As shown in table 1, digitization has had a rapid and significant impact on the existing reformatting workloads of preservation departments. Within the five years of digitization statistics gathered, the number of individual sheets digitized surpasses the number of sheets photocopied by a small margin (13,340). A more drastic comparison can be made by the number of volumes digitized (51,254) to the number of volumes photocopied (20,112). Preservation photocopying has typically been the reformatting option of choice when an actively used volume is found damaged in the stacks. Is this traditional process being replaced with a digital alternative?

In examining digital-reformatting tasks over the past five years, we see that the number of items completed does not steadily increase across years as one might expect. (See table 2.) This incon-

sistency over five years may be explained by a time-sensitive grant for a specific amount of reformatting (as suspected in the large spike in 2000 for sheets digitized). The most notable aspect of table 2 is that even in the first year ever of digitiza-

TABLE 2 Average Digital Reformatting Tasks over Five Years					
	1998	1999	2000	2001	2002
Volumes digitized	615	318	986	347	582
Sheets digitized	1,517	1,900	22,151	2,856	2,599
Nonpaper items digitized	2,086	1,969	1,524	2,676	2,762

tion statistics gathered, the number of items completed is quite high. This suggests that digitization took hold quickly within preservation departments as a reformatting task and, on average, has simply increased the workload completed there. In a 2001 survey, preservation administrators noted that the most significant change over the past five years in preservation departments was the inclusion of digitization tasks.[3]

How does the number of items reformatted via digitization compare with the number of items reformatted via traditional processes such as photocopying and microfilming? Figure 1 shows that digital tasks account for ten percent of the total items processed in a preservation department over the past five years.

One would assume that in order to maintain a steady workload, the number of items processed digitally would increase and the number of items processed with analog methods would decrease. This interaction was tested in the statistical package SPSS, with the expectation that the numbers of digital and analog reformatting tasks would be negatively correlated. A two-tailed partial correlation was run, testing items reformatted digitally against items reformatted using analog methods, controlling for year. The statistical test shows that the two processes are not negatively correlated, demonstrating that as one process (digital reformatting) grows, the other (analog reformatting) does not decrease accordingly. This suggests that the total items processed in preservation

FIGURE 1
Number of Items Reformatted with Digital and Analog Means, 1998-2002

10%
Digital reformatting
(807,959 items)

90%

Analog reformatting
(7,215,754 items)
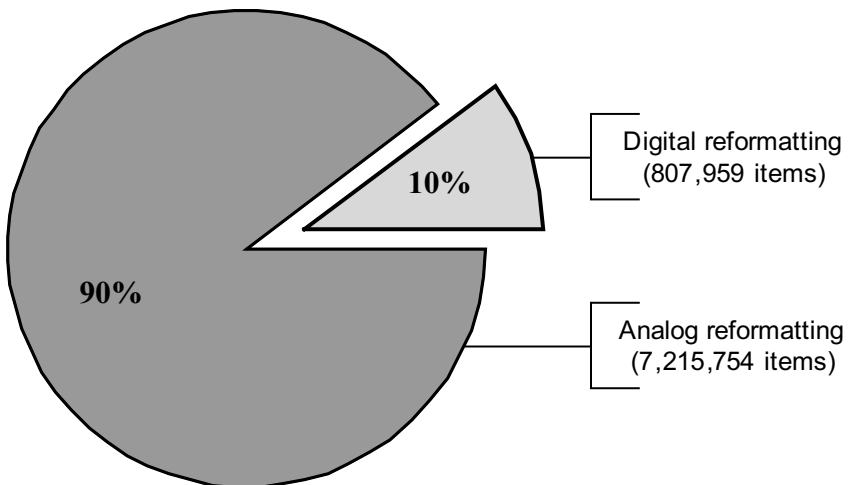
departments has not remained steady but, instead, has increased by ten percent over the past five years due to the additional reporting of digitization tasks.

ARL Staffing Statistics

As the workload for preservation departments has increased due to the new digitization tasks added, one would expect that staffing would increase accordingly. However, according to ARL data, this is not the case. As demonstrated in figure 2, staffing at all levels (professional, paraprofessional/support staff, student) has remained steady over the past five years.

Professional staff shows a slight increase overall, averaging 4.98 people in 1998 and 5.21 in 2002. The number of support staff shows a slight rise as well, averaging 10.83 people in 1998 and 10.93 in 2002. The number of student workers also rises slightly but then decreases, averaging 2.29 people in 1998 and 1.90 in 2002. (See table 3 for averages across five years.)

This look at staffing statistics demonstrates that preservation departments are doing more work with about the same number of employees. Answers to questions 7 and 10 from the Mohlenrich survey

of preservation administrators show a slight loss in existing staff performing the reformatting tasks of photocopying and microfilming and a slight gain in existing staff performing digitization tasks such as selection, metadata creation, and quality control.[4] This loss/gain does not affect a significant staffing decrease/increase, suggesting that the shift in staff within departments from traditional tasks to digital tasks is to try to accommodate the increase in workload.

The Merrill-Oldham model of mature libraries proposes a benchmark for the number of personnel appropriate for Group 4 libraries.[5] It states that this group should have more than seven professional staff and more than twenty support staff (including student workers) for a total of more than thirty total staff. As evident in table 3, not one year of staffing in Group 4 reaches the proposed benchmark. In the Mohlenrich survey, (2001) 25 percent of library administrators noted that they did not feel that they were meeting their library's needs due to understaffing and 29 percent of administrators noted that personnel was a factor that would significantly influence the direction of their preservation program in the future.[6]

FIGURE 2
Average Number of Group 4 Preservation Department Staff, Across Five Years
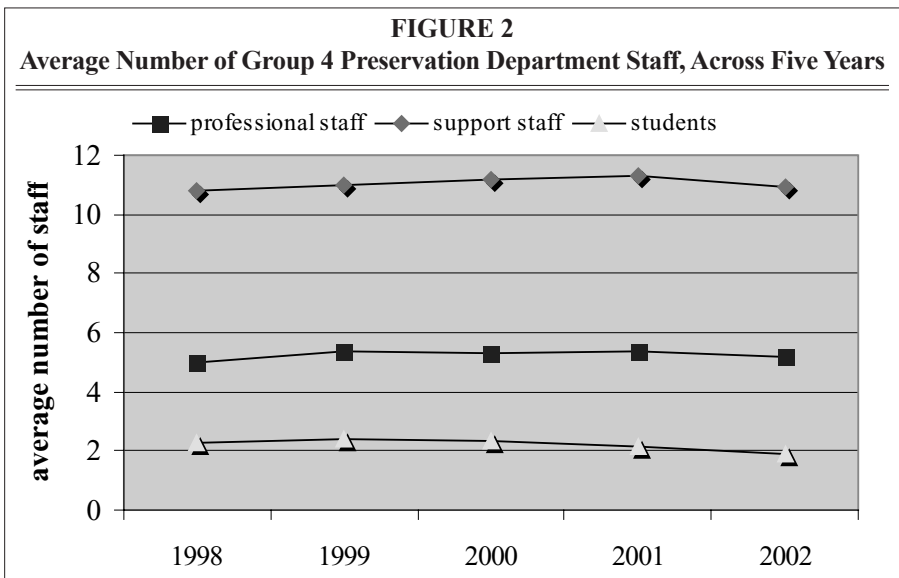
Administration of Digitization in Preservation Departments

Does the inclusion of digital tasks in a preservation department demand that preservation administrators have skills in those areas? To answer this question led to looking at the required qualifications of open job calls for preservation administrators as posted on the Preservation Administration Discussion Group List archive. In 2004, four job descriptions specified a desire for expertise in digitization.

In one of the job postings, the position of preservation librarian was to be in charge of several units, including reformatting, which performed "digital imaging activities."[7] Another entry-level position noted that "experience, … including digitization" was a preferred qualification.[8] One university sought two digitization experts: one as the head of the preservation department, who would supervise a conservation librarian "dedicated to preservation imaging and reformatting," and the other to specifically manage the preservation reformatting program, "providing leadership with respect to the preservation of digital collections."[9] From these job descriptions, it is evident that digitization goals vary widely across universities, requiring preservation administrator candidates to self-select the types of institutions in which they seek employment, according to their skills.

This brief examination of recent job postings for preservation administrators can offer only a snapshot of requirements over the past calendar year and does not suggest that these qualifications for digital expertise are new. This look at the postings simply points out that the current state of preservation administration requires some level of digital expertise.

A study proposed by the University of Pittsburgh suggests that digital expertise cannot be gained via traditional library programs. This study on the preservation education needs for information professionals notes that digitization should be a key curricular component, in addition to skilled conservation for alternative media such as DVDs, maps, and architectural drawings.[10] The University of Pittsburgh proposal notes that access to, and education about, the preservation of digital collections is not adequately addressed within a traditional library science curriculum. The study proposes to reevaluate the objectives of preservation education in order to guide the studies of future information professionals.

Intralibrary Collaboration

A strong preservation department collaborates with other library staff to accomplish preservation tasks. ARL has recognized the intralibrary efforts of preservation departments by collecting statistics that address preservation tasks done within a library, but outside the preservation department. ARL permits each reporting institution to decide what it considers a "preservation activity," but one can assume that preservation activities range from simple activities such as flagging a damaged spine, missing pages, or inclusions to complex activities such as assisting in collection management or treatment decisions. A Group 4 library, The University of North Carolina (UNC), exhibits this type of effective collaboration within its library system.

The numbers of total UNC library staff performing preservation tasks (as viewed in table 4) demonstrate the strength of UNC's intralibrary collaboration. Although UNC maintains significantly fewer preservation department staff

	TABLE 3 Average Number of Workers per Staff Level of Group 4 Libraries, across Five Years				
	1998	**1999**	**2000**	**2001**	**2002**
Professionals	4.98	5.38	5.31	5.36	5.21
Staff	10.83	10.99	11.16	11.33	10.93
Students	2.29	2.40	2.36	2.15	1.90

members than the Group 4 average, it counters this by promoting preservation tasks throughout the library, encouraging other staff to participate in the overall preservation of the collection. The result of this proactivity is that the total UNC library staff performing preservation tasks is very similar to the average number of staff performing preservation tasks across all Group 4 libraries.

Testing the Group 4 Model against Actual Digitization Practices

The framework used for this research includes the author's modernization of the Group 4 model to include digital-reformatting tasks as part of the photo-duplicating unit. The re-imagined model assumes that the tasks are performed in-house, unlike the other groups of the Merrill-Oldham model that outsource such tasks. To test the appropriateness of this framework, the author pursued conversation with six of these departments to determine whether they do, in fact, participate in digital-reformatting tasks. The six preservation departments contacted are Michigan, Yale, Cornell, Stanford, Harvard, and California-Berkeley.

In addition to confirming whether these preservation departments perform digital-reformatting tasks, the author asked whether their libraries considered the tasks to be "preservation." The author also inquired whether the preservation departments performed all digitization in-house or outsourced some of the tasks. Those who outsourced some digitization tasks were asked what percentage of total

digitization tasks were outsourced and whether a particular type of digitization task was vended out consistently (e.g., no in-house capacity for audio or video reformatting). Finally, the author requested the number of preservation staff committed to digitization tasks.

In answering whether their preservation departments perform digital-reformatting tasks, all responded positively except for Stanford. Due to a recent reorganization, all of Stanford's digital activities are now housed at the university level and report to an academic group outside the libraries. As a result, the preservation department is no longer involved in digital production or reformatting. Stanford has gathered its digital expertise into a discreet group outside the library, called the Digital Services Group. The other five responding departments noted that their libraries consider their involvement with digitization to be "preservation," but determining when a digital item has been "preserved" remains an outstanding issue that will be touched on in a subsequent section on "quality" in this paper.

Only one of the five preservation departments actively involved in digitization tasks performs these duties exclusively in-house. The others outsource some of their digital-reformatting tasks at a rate from 25 to 100 percent. The type of task outsourced is not consistent across the five institutions and includes such tasks as slide or microfilm scanning, OCR, and metadata construction. Four of the five stated specifically that their departments were involved in the prepa-

TABLE 4
Total Number of UNC Staff Performing Preservation Tasks across the Library, Compared with the Group 4 Average

	1998	1999	2000	2001	2002
Nonpreservation staff	11.76	13.02	21.93	27.90	21.13
Preservation staff	10.65	8.30	6.65	6.50	10
Total UNC staff performing preservation tasks	22.41	21.32	28.58	34.40	31.13
Group 4 average staff performing preservation tasks (preservation and non-preservation staff)	33.64	30.58	30.30	31.43	31.13

ration of materials and specifications for outsourcing and/or quality control upon their return.

Identifying the number of staff involved in digital-reformatting tasks proved to be a difficult question to answer due to the areas of proficiency required to perform the tasks. Because the development of digital projects draws on a wide area of expertise, from deciding which items to reformat to performing appropriate bibliographic information for the resulting digital file, most noted that they utilized skills from staff in other departments within the library or university. This consultation, though meaningful to the success of a digitization task, is difficult and time-consuming to quantify. The number of staff dedicated to digitization, therefore, ranges from .15 FTE to 6.5 FTEs at five of the departments contacted.

The resulting conversations from contact with six of the preservation departments described as members of a reimagined Group 4 suggests that, although five of them do perform digitization tasks, their workload and staffing are varied. The results of conversation with six of the eighteen libraries suggest that the data supplied for the ARL preservation statistics should act as only a starting point for cross-library comparison and benchmarking.

What the Statistics Do Not Tell Us

It is understood that, as new statistics are gathered, it takes several years to interpret what the library is requested to report, to determine how to best represent the workload via statistics, and to measure efforts against similar libraries. When ARL adds new statistical measures, it maintains a period during which the reporting is optional and then becomes mandatory. The years on which this analysis is based can be described as the "learning years," or the years during which preservation departments come to understand how to report statistics for their digitally reformatted items. During this learning period, the ARL makes the statistical reporting an optional process, encouraging discussion

within member libraries on how to determine each group's reporting methods. Therefore, the analysis presented here can only be suggestive of trends, rather than definitive. The incomplete data offered during the five-year learning period prompt discussion of how the questions in the annual survey are to be interpreted and what is to be gained by reporting those numbers.

Quality

As they are gathered in the proposed statistical measures, the data may never tell us about the quality or depth of the work produced digitally. The annual survey itself provides some guidance about how to count an item reformatted using digital means, but it does not define the expected quality of the product. In contrast, clear guidelines are offered on how to determine whether an item has been photocopied or microfilmed appropriately, citing national standards. No such guideline is offered for digital reformatting. Instead, the survey offers a guideline on how to define what it means to "digitiz[e] for preservation purposes."[11] The guideline is meant to assist member libraries in determining how to count items reformatted digitally. What is not evident, however, is a link to a standard or schema for how to produce digitally preserved items. It may be understood that something reformatted using digital means may be considered "preserved" if scanned/photographed at 600 ppi, saved as a TIFF file (uncompressed), and given appropriate metadata for future resource discovery, but this is not explicitly stated in the survey. Therefore, this leaves open the interpretation of what "digitiz[ed] for preservation purposes" means. This ambiguity may prove to be problematic in the future when attempting to evaluate digitization statistics across ARL member libraries.

Time

Assuming that at some point a standard is adopted and linked to the annual ARL preservation statistics outlining the nec-

essary characteristics of a digital file to be considered a preservation copy, the issue of time may still remain an issue. At this point, a preservation department may digitally reformat an item in an appropriate size and format and store it with a commitment to longevity/forward migration. This library would count this effort as one item digitally reformatted. Another library may take the same first step but add on additional processing, such as thumbnail creation, metadata creation, or additional cataloging. This effort would still be counted as one item digitally reformatted. This disparity in time given to digital reformatting might be addressed through the existing schema of conservation treatment statistics.

Already existing in ARL statistics are levels of conservation treatments, ranging from less than fifteen minutes (level 1) to two hours or more (level 3). These time lines of efforts given to treat an item also may be suitable for digital reformatting. Level 1, the minimum amount of effort/time, may be appropriate for an item that has been simply scanned as a master file. Level 2 may be appropriate for reformatting as a master file, in addition to the creation of a smaller-sized file (for low-resolution reproduction) or thumbnail (for viewing in an OPAC). Level 3 may represent the time needed to create a master file, a lower-resolution file, and complete metadata, to be used for future resource discovery.

Depth of Curatorial Effort
In addition to the quality of the digital file and the time needed to create it is the effort of preservation department staff to assist in choosing which items to reformat digitally. Already in the infrastructure of a library are procedures for identifying when an item should be reformatted due to deterioration; guidelines are in place for when to photocopy or microfilm. This process is complicated with the inclusion of digitization because the items are not simply reformatted and put back on the shelf. Additional infrastructure is re-

quired as an item is digitally reformatted because the item must be put back on a "digital" shelf, using a schema that makes the item retrievable by a library patron. This new step includes preservation staff in collection management decisions, helping to determine whether the item makes sense for inclusion in the library's digital collection when it may have existed previously in the analog collection.

Deciding what should be held in a digital collection requires a different level of commitment than that which is held in an analog collection. A book generally remains a book on a shelf, but a digital file is alive as it migrates to future format(s). Understanding that maintaining the use of these files, therefore, requires a deep curatorial commitment on the part of collection management and preservation departments. This type of effort cannot be gathered in any statistic but, rather, will be reflected in the appropriateness of what is digitized through the use of those files by patrons.

As ARL statistics are gathered now, a single statistic may not represent the true efforts given to reformatting an item for preservation purposes. The true range of digitization practice may not be appropriately captured by a single number because "digitization" is still too loosely interpreted to be useful in statistics such as those proposed by ARL. The amount of time is also not currently addressed in the statistics but may be easily addressed through the creation of a schema similar to that of conservation treatment. Quantifying deep curatorial efforts versus a "scan everything" approach may never be appropriate for consideration in statistical form.

Discussion
Though exciting digital projects are being created exponentially in preservation departments across the ARL membership, it is clear that digitization has not replaced traditional preservation strategies. The author expected a decline in the number of items reformatted using analog means

that has not occurred as the number of digitally reformatted items increases. This suggests that digitization is not taking the place of preservation photocopying or microfilming. Instead, it is a separate process being performed in some preservation departments, increasing the number of items processed there by ten percent. Staffing has not yet increased to meet this dramatic shift in workload.

It is clear from these analyses that quantifying a concrete number of preservation department staff performing digital tasks is difficult due to the variability of the tasks. Some digital-reformatting tasks require staff to seek expertise and consultation outside the department, such as deciding which items to retain after conversion, how to construct metadata, and how to scale images for presentation in a library OPAC. Conversely, the same staff that seek expertise in an area can act as digital consultants to university departments outside preservation, providing their in-house standards to departments considering their own digital projects, offering guidance on how to store digi-

tal files, and contributing to a national discussion of digitization standards for preservation. This fluidity of task assignment and expertise/resource sharing suggests that quantifying the number of staff involved in the reformatting of one digital item may be difficult until libraries determine consistent work flows and workloads related to digitization.

Conclusion

The ARL recognizes the impact that digitization is having on preservation departments by requesting data from member libraries over the past five years. The data show that digitization is having a significant impact on the workload performed there but as yet has had little impact on the staffing numbers in those departments. As digitization tasks are tracked over subsequent years, patterns of the effects on workload and staffing will emerge. Continued tracking of similar variables over the next five years also may be more revealing and suggestive of long-term impact of digitization on preservation departments.

Notes

1. Association for Research Libraries. "ARL Member Libraries," 2004. Available online at http://www.arl.org/members.html. [Accessed 23 May 2005].

2. Jan Merrill-Oldham, Carolyn Clark Morrow, and Mark Roosa, *Preservation Program Models: A Study Project and Report* (Washington, D.C.: Association of Research Libraries Committee on Preservation of Research Library Materials, 1991).

3. Janice Mohlhenrich, *Preservation and Digitization in ARL Libraries: A SPEC Kit* (Washington, D.C.: Association of Research Libraries Office of Leadership and Management Services, 2001).

4. Ibid.

5. Merrill-Oldham, Morrow, and Roosa, *Preservation Program Models.*

6. Mohlhenrich, *Preservation and Digitization in ARL Libraries.*

7. The Preservation Administration Discussion Group List Archives, 2084. Available online at http://palimpsest.stanford.edu/byform/mailing-lists/padg/. [Accessed 23 May 2005].

8. Ibid., 165.

9. Ibid., 169, 2337.

10. University of Pittsburgh, *Preservation Education Needs for the Next Generation of Information Professionals*. Available online at http://www.sis.pitt.edu/~kgracy/Pres_Edu_Study.htm. [Accessed 23 May 2005].

11. Mark Young and Martha Kyrillidou, eds. *ARL Preservation Statistics 2001–2002: A Compilation of Statistics from the Members of the Association of Research Libraries* (Washington, D.C.: ARL, 2003).