Scarce and Endangered Works: Using Network-level Holdings Data in Preservation Decision-making and Stewardship of the Printed Record

Jacob Nadal (Executive Director, ReCAP: The Research Collections and Preservation Consortium), Annie Peterson (Preservation Librarian, Tulane University), Dawn Aveline (Preservation Officer, UCLA)

Abstract

We propose methods of making preservation decisions based on holdings data for library collections, rather than on intrinsic characteristics of the damaged items or the preferences of local collection managers. This study focuses on monographic collections, but it builds on related work in serials preservation and we argue that there is potential for a common framework that encompasses both types of published materials. These methods borrow from economic assessments of risk and simple–to-use readily available data in a way that encourages fast and predictable decision-making. The outcome is a risk-management framework focused on ensuring access to a complete copy of a work at a given point well into the future, without requiring a specific set of future conditions for their success. Instead, these approaches are designed to serve as an adaptive framework for ensuring access to the records and artifacts that are in the library’s care. Our work evaluates the role of regional and consortial holdings, as well as the relationship of print to digitized holdings in managing preservation risks. We also describe the methods that are used to limit undue risk and allow input from various stakeholder groups.

Introduction

Beginning in 2008, the University of California, Los Angeles (UCLA) Library began a series of experiments and pilot projects to explore the issues at stake in stewardship of the...
printed record and the best ways to structure preservation programs to support the Library’s commitments. The “preservation review” function developed for UCLA Library uses holdings data to expedite preservation responses to critically damaged or decaying materials, as soon as they are identified in the course of day-to-day library operations. Based on the results garnered from UCLA’s preservation review process, we propose methods for preservation management in a cooperative but not centrally administered library environment.

There are two broad outcomes from preservation review: withdrawal or replacement. Replacement options include reformatting to a digital surrogate or print facsimile (or both), or the purchase of another copy in good condition. If a copy in serviceable condition can be found on the secondary market, we purchase it as a replacement and withdraw the copy in preservation review. When copies are not available for purchase, we send the item to the Internet Archive for reformatting, followed by digital deposit to the HathiTrust, prior to withdrawal or deposit in the Southern Regional Library Facility (SRLF), a closed stack preservation repository.

Decisions about the course of treatment are primarily made by assessing the WorldCat holdings data related to the item in question. We default to replacement for items with fewer than twelve (12) global holdings, fewer than three (3) California holdings, and zero (0) additional University of California (UC) system-wide holdings; withdrawal is the default for items with more than twenty-six (26) global holdings, more than five (5) California holdings, and more than two (2) UC holdings. These criteria are inclusive; the volumes must meet all of the criteria to receive a particular outcome. Collection managers review all other items and decide between replacement and withdrawal. These thresholds allow us to estimate the probability that, after taking the local preservation action, a complete
copy can be found or reconstructed in 2100 CE. For our standard scenario, this probability runs to approximately 78% within California, and better than 99.5% across global holdings.

The first section of this case study describes the process of collecting and evaluating holdings data and applying these results to preservation decision-making. The second part extrapolates from the lessons learned in our preservation work to issues implicit in collaborative print archiving projects and other types of network-level collection management. In both sections, we confront the problems inherent in making local operational decisions in the face of incomplete or imperfect information. This paper expands upon a white paper published by Peterson and Nadal in 2011, updating that work to include additional data from UCLA along with Aveline’s observations about the development of the preservation review program.¹

**Background and Literature Review**

Large-scale digitization efforts in research libraries have progressed rapidly. By July of 2014, the HathiTrust contained over 11 million volumes, a quantity comparable to any of the world’s major collections of record.² According to a recent report, however, only 27% of the then 10 million volumes in HathiTrust were in the public domain.³ The existence of this body of digitized materials does not immediately create ubiquitous access to the pre-twenty-first century printed record. Numerous factors remain at play to determine the future methods of providing access to library collections: intellectual property, licensing, and copyright; the applicability of fair-use standards and first-sale doctrine to reformatted and born-digital works; the role of libraries in material culture scholarship; the diverse and changing field of user-preferences for print and digital versions of a work. The list can go on, but whatever the near-term obstacles and unresolved questions, the intention of the library community is clear:
the effort is well underway to create ubiquitous digital access to the published record. In turn, the print collection is being positioned as an adjunct or interdependent part of the library. Printed books meet a reading preference, function as a source of artifactual evidence, serve as a backup in case of loss to digital collections, and provide a future source for re-mastering to create new digital objects.4

In general, preservation decision-making, cost management, and selection for preservation have been discussed from the perspective of how local decisions were made to manage capacity or to argue the need and value of libraries investing in preservation efforts overall. The national microfilming effort that began with the US Newspaper Program in the 1980s provided the impetus for much of this work. Robert Hayes’ 1987 report, The Magnitude, Costs and Benefits of the Preservation of Brittle Books, identifies keys factors in selection and costs associated with the microfilm projects into the 1980s and will be used as a source for cost-comparisons in this paper.5 Michael Lesk’s “Selection for Preservation of Research Library Materials” provides a thorough overview of selection methods for preservation effort.6 The focus of selection for reformatting in the late twentieth century was on pre-coordinated sets of materials, drawn from the areas of the collection identified as likely to yield preservation problems. These sets were generated through a “great collections” approach or through consultation of a scholarly bibliography. A great collections approach “begins with an evaluation of the relative strength of collections by classification” to produce “a ranking of comprehensiveness of various libraries in specific subject areas.”7 Scholarly lists are to be taken from published bibliographies or through consultation with scholarly societies or ad hoc academic committees. Once a list is established or a great collection selected, preservation reformatting proceeds within that collection.
In his overview of the preservation selection methods that were becoming established in the 1990s, Robert Mareck describes a two-part process of selection for preservation and selection in preservation. The first part considers the decision about whether a particular library ought to preserve something, the second a decision about how to preserve the item in question. Implicitly, this process represents a shift from a collection-driven strategy, one that declares the value of a set of library resources in advance of an item-driven strategy. Mareck describes his model in terms of a “macro preservation decision” that “is not based on the individual discrete item, but on its inclusion in a larger coherent body of material.” Despite the language of macro decisions and bodies of material, however, the critical distinction must be made between two broad types of library decision-making processes. Item-driven approaches originate from an item-in-hand and then investigate relevant context. Collection-driven approaches begin with a hypothetically complete corpus and then investigate whether the specific members of that corpus exist in a given location and in serviceable condition.

Mareck provides a four-part model for connecting the critical, or “macro,” assessments with the technical, or “micro,” assessments in his model. Although Mareck presents a structured container for the critical and technical decisions, and allowing that he is summarizing practices, his description of the “macro” stage remains loosely structured at best. He provides a list of questions to guide decision-making, but without any metrics for determining an outcome or indication of what specific data would prompt a decisively positive or negative decision within any set of questions. Mareck’s second stage, the “micro preservation decision,” is more focused, and consists of a set of advantage and disadvantage statements for each standard preservation outcome. Our process allows for more efficient
decision-making by establishing clear metrics for the “macro” decision-making, and by using some of the data for “macro” decisions to guide “micro” decisions, as well.

Surveys and statistical approaches to preservation decision-making trace their roots to the Yale condition survey of the 1980s\textsuperscript{10} and to the survey strategies that Carl Drott developed for collection management beginning in the 1960s.\textsuperscript{11} These surveys have primarily focused on understanding the physical condition of library collections as a guide to developing preservation strategies. Twenty-first century preservation efforts have shifted towards a balanced strategy of artifact conservation, digital reformatting, and the development of cooperative repositories for long-term storage. Much of the earlier literature has lost its currency due to simple changes in technology or more complex changes in the preservation decision-making environment. Jan Merrill-Oldham in “Taking Care,” in the \textit{Journal of Library Administration} offers a clear summary of the current context for preservation decision-making.\textsuperscript{12} Oldham reviews the primary arenas of preservation activity and discusses the primary factors at play in each. In this survey article, however, Oldham does not investigate specific methodologies for decision-making, selection, or specifics of cost assessment.

Julie Mosbo and John Ballestro write about the feasibility of purchasing from secondary sellers, finding that purchases off the used, secondary, and out-of-print markets meet the needs of the Southern Illinois University Morris Library at a net cost-savings.\textsuperscript{13} Mosbo and Ballestro found that only 3.5% of the items ordered required preservation treatment, but these were also works printed within the last decade. Very little has been written on the use of replacement as a preservation strategy specifically.
Libraries have a substantial history of cooperating to store the published record, through formal efforts such as the Center for Research Libraries, and through the uncoordinated network effect, where numerous research libraries across the world purchase copies to serve their local users but also share scarce resources via interlibrary loan. Cooperative print-archiving projects developed in the first decades of the twenty-first century as a corollary to large-scale digitization of print collections and the broad shift in usage of serials from print to electronic versions. On July 21, 2003, the Center for Research Libraries convened *Preserving America’s Printed Resources (PAPR)*, a conference focused on “the role of repositories, depositories, and libraries of record.”¹⁴ This conference considered the major issues around the mission of libraries in cooperative preservation as well as key operation concerns implicit in collaborations around shared resources.

Just over a year after the *PAPR* conference, in October 2004, Google launched the Google Books Libraries Project followed shortly after by the Open Content Alliance and Microsoft LiveSearch collaboration.¹⁵ The advent of large-scale digitization placed a new focus on the role of print in library collections and raised questions about the relevancy of previous methods of selecting candidates for preservation reformatting. These trends are described in Oya Reiger’s 2008 report, *Preservation in the Age of Large-Scale Digitization.*¹⁶

Late twentieth century preservation reformatting efforts selected damaged items from within the larger collection. The large-scale digitization projects of the twenty-first century do the opposite. To create a workflow with high and predictable throughput, these digitization efforts reject damaged items from a collection– or library–wide workflow that is not strictly constrained by subject parameters or scholarly input. Content reformatting, a desire for text in digital form, drives contemporary reformatting practices. The systematic
content digitization projects of the early 2000s do not provide the same primary benefits as preservation reformatting projects, since they frequently exclude the most critically damaged artifacts. While preservation is not the primary goal of these projects, they do offer an ancillary benefit, since more copies in more formats spreads risk across a more diverse set of formats with different advantages and disadvantages.

In 2009, Ithaka S+R released *What to Withdraw? Print Collections Management in the Wake of Digitization*. This report encapsulates the lessons learned from the JSTOR project, digitization projects with Google and the Internet Archive, and the work conducted on print archives during the first decade of the twenty-first century. The report and an accompanying Excel-based calculator introduce a methodology for deciding which print journals libraries can withdraw responsibly and how that set of materials can be expanded to allow libraries the maximum possible flexibility in managing their collections. *What to Withdraw* builds on a study that Ithaka commissioned from Candace Arai Yano, an operations research professor at University of California, Berkeley, and her colleagues in the Department of Industrial Engineering and Operations Research. Yano’s study has also been used as the fundamental building block for the decision-making process described in this case study.

**Methodology I: Risk Assessment and Replacement Model**

Yano’s framework considers time and rate of loss for serials, based on data from the JSTOR validation process, library claims rates filed with insurance companies, and loss rates for circulating materials. These data are used in a calculation for estimating the number of required initial items to produce a single perfect copy at a given point in the future. This method of assessing survival probabilities is similar to methods used to resolve supply chain
management issues around the production of spare parts for manufactured goods or to do risk modeling for redundant array of inexpensive disk (RAID) systems.\textsuperscript{19}

Yano assumes that the intention for this final copy is perfect information content, not necessarily a single perfect artifact. Because her study focuses on journals, hybridization of a complete work from multiple sources is feasible as part of a later digitization process. More specifically, her work assumes that “[v]irtually all of the journals will be stored in the form of bound volumes, so we take a bound volume… as the unit of analysis.”\textsuperscript{20} Applying this methodology to monographs raises the question of whether they exhibit the same survival characteristics as journals.

If one considers the problem as merely the loss of bound volumes, then serials and monographs should be interchangeable. The performance of book structures bears no necessary relationship to the intellectual content of the work. However, the process of loss and reconstruction is more precisely understood as the loss of some knowable percentage of a serial, itself composed of bound volumes, over time. The extant volumes from one instance of a serial replace the lost elements from other instances of the serial, each of which has also suffered loss. The survival probability represents a measurement of the likelihood that all of the extant series have not lost the same parts and thus eliminated the possibility of creating a perfect hybrid series. Using this model for monographic assessment becomes intuitively problematic. For clarity, the crucial question should be framed around the notion of repetition of parts in a series. The mathematics remain agnostic about bibliographic conventions, so intellectual series in and of themselves yield to the relationship of parts to wholes, or members to sets.
Although monographs are not explicitly cataloged as continuing resources, they are easily conceptualized into a variety of sets, based on common points of bibliographic identity, such as their author, year of publication, or language. Individual monographs can also be treated simply as members in a set of all instances of that monograph. Subsets could be conceived of in a variety of ways, including geographical regions and consortia, which are already common points of identity and collaboration among libraries.

In this vein, Yano’s work draws on Martin L. Weitzman’s “The Noah’s Ark Problem,” which uses a library model to evaluate biodiversity risks by “conceptualizing the underlying conservation unit – the ‘species’ – as if it were a ‘library.’”21 Weitzman argues that “[t]he book collections in various libraries may overlap to some degree” and develops a measurement of diversity as “the size of the set that consists of the union of all the different books in all the libraries” in the study.22

Our method follows Weitzman by treating WorldCat as our assemblage of libraries and UCLA Library’s collections as one branch within the larger set. Weitzman’s model proposes a diversity function, $V$, for a set of libraries, $S$. Each library within $S$ holds a set of books held in common with other libraries, $J$, and a set of books unique to that library, $E_{library}$. Our process examines each book in preservation review to determine if it belongs to the set $E_{UCLA}$ or $J$, and if the latter, to decide if its withdrawal adversely affects $V_{(global)}$, the diversity function of libraries globally. Because we are working with real libraries rather than the abstract library that Weitzman models, we investigate $E_{UCLA}$, $E_{UC}$, and $E_{CA}$, in addition to $J$. Furthermore, for all of the theoretical value of Weitzman’s model, it is Yano’s formulation that does the heavy lifting by allowing us to quantify the extent of risk to the shared collection, the $J$ set of $V_{(global)}$, created by withdrawal or replacement.
The question at issue, “of the set of items, what is the chance of loss”, can be meaningful whether the items are individual bound volumes (monographs), or sets of bound volumes (serials). The issue that remains is determining the loss rate for monographs, in case it is different than the loss rate for serials. Supposing 100 instances of a serial with 100 volumes, 1% loss means the loss of one volume. In the monographic parallel, something like 100 copies of a monograph with 100 pages, 1% loss means one page. Although the same risk evaluation formulas can be applied, monographs may seem more fragile than serials, since 1% loss occurs at a dramatically lower threshold. To account for this in our analyses below, we use loss rates derived from observation and we double that rate to seek thresholds that are safe for both. We also use several sets of monographs (UCLA, UC, CA, and Global), in effect creating series at different consortium and regional levels, imitating the pattern of serials.

Methodology II: Data Collection, Accuracy, and Applicability

For our study, we used data from 1,408 items selected at random from among the brittle materials identified by collection managers and at circulation points during the last decade at UCLA Library. This provided us with a set of materials from across the Library’s entire collection that is large enough to be statistically useful, provided it is truly random and free from major biases in selection. We would ideally draw on a random sample of damaged materials from the entire collection to develop policies, strategies, and preservation actions at the institutional level, but there are significant barriers to the discovery of a sufficient sample of these materials.

The common traits of interest among preservation candidates—paper decay and mechanical failure of book structures—are caused by inherent factors as well as external
influences. Neither set of causes is tracked in library databases, so automated generation of a randomized sampling list is impossible. The only readily available proxies for severe damage are publication date and perhaps country of publication, but these factors show low correlation to severe damage: although severely damaged materials are somewhat likely to be old, older materials are not as likely to be severely damaged. Locating an ideal random sample therefore requires surveying at the shelf across the entire collection. For UCLA Library, with a collection of over 9 million items, the walking distance of the survey alone could approach 40 miles.

The sample used for this study matched the profile of severely damaged materials as they appear in our general preservation collection surveys. The general surveys include materials from all UCLA Library branches and collection areas, suggesting that this sample is representative of the collections as a whole. Because the preservation department is a fairly new entity in the UCLA Library, no systematic plan has been in place for collecting or treating brittle materials, so we do not believe that our sample is biased by any long-running and consistent past practice for collecting candidate volumes for preservation treatment.

**Preservation Treatment Outcomes**

Preservation options for severely damaged materials are limited. These items are characterized by extreme paper decay, severe structural damage, or both. As a consequence of paper decay, the materials used for repair often exceed the mechanical strength of the paper. The double-fold test characterized this problem by counting how many times a page could fold before breaking apart. A sheet of paper in good condition will survive dozens of folds at least; hundreds if one has the patience. The books with paper decay in this study often fail before one fold is complete, or at best, two or three. Even the small level of
adhesion generated by a repositionable note (such as the Post-it brand from 3M), which is on the order of .1 – 2.7 N/25mm, suffices to peel away a layer of paper from many of the items we considered. Structural damage to the case or text block of the volume forms the other main factor. The materials in this study included some that the dog literally ate; most others exhibited enough damage that conventional repair or commercial re-binding was not possible. In many instances, chemical decay and structural damage play upon one another to create items with complex, compound problems. The physical integrity of all items subjected to preservation review has been so severely compromised that the item can no longer serve its function as an information carrier, and cannot be circulated without risk of further damage or content loss.

In this study, preservation review leads to two broad outcomes: withdrawal or replacement. Replacement options include reformatting to a digital surrogate, print facsimile (or both), or purchase of another copy of the same edition in acceptable condition. The final disposition of the original volume is implicitly withdrawal following reformatting or replacement. If items of substantial artifactual value are identified for replacement and another print copy cannot be substituted, we reformat the item as it came to preservation review, place the original artifact in a suitable enclosure, and deposit it into closed storage with an in-house use only status. In principle, we also reserve the option of heroic conservation efforts should materials of dramatic artifactual or associative value come to light, but this has not occurred in practice.

Yano’s survival probability algorithm assumes a survival scenario for a hybrid print and digital information environment where i) the availability of digital version will lead to low print usage, ii) the movement of print materials from open stacks to closed depositories
will reduce loss and damage, and iii) reduce chemical decay through environmental controls. The treatment outcomes developed for our preservation review process are designed to achieve these conditions through digitization of scarcely held materials and, when replacement copies are purchased or a retention decision is made for the original book, relocating it to the Southern Regional Library Facility (SRLF). SRLF deposit with a restricted use status performs the functional role of a dark archive in Yano’s model, with good environmental conditions, near-zero loss, and an option for later reformatting. When the replacement option leads to the discarding of a print item, it is only after digitization. The loss to the library system is minimal in these instances. The materials in question are severely damaged, making them poor exemplars of the artifactual value of that work. Creating a digital surrogate allows the library system to gain the added preservation benefits of working with a hybrid environment and spread the risks to the survival of content across multiple formats.

**Holdings-Based Decision-Making Scenarios**

In developing this project, we speculated that complex decision-making strategies could be replaced with simple holdings analysis. To the extent that our simplified strategy is successful, the total cost of preservation actions can be more closely evaluated and controlled. In addition, the otherwise diffuse decision-making efforts of library staff can be clearly located in the organization, creating an opportunity for better accountability, reduced costs, and fewer delays in service. To test our assumption that holdings-based decision-making could be an effective proxy for other means of preservation decision-making, we created three decision-making formulas. We describe these holdings-based scenarios as aggressive,
moderate, and conservative, in reference to the degree to which they lean toward the “aggressive” option of disposal of materials instead of replacement, repair, or reformatting.

Each of the three formulas has two thresholds, leading to three classes of preservation decisions. The first is the replacement threshold, the number of holdings in WorldCat below which we default to taking preservation action. The total number of WorldCat holdings includes only items of the same edition and format. Below this number, the default decision is to replace or reformat. The second threshold, withdrawal, marks the level of holdings above which we default to withdrawal of our copy.

We selected replacement thresholds to create specific survival probabilities based on Yano’s model, likewise for the upper thresholds in the aggressive and conservative formulas. The withdrawal threshold in the moderate formula reflects local practice, and was derived from an analysis of past collection management decisions at the UCLA Library. These decisions showed a strong trend towards withdrawal when global holdings in OCLC WorldCat were greater than twenty-six (26), so we selected this number as the upper boundary for the moderate scenario. The UCLA Library Preservation Department currently follows the moderate formula. Finally, we implemented a policy decision to always take preservation action when UCLA holds the last copy in the University of California system or one of only two copies in the state of California, independent of other holdings data. Although almost an afterthought in our early investigations, this policy has turned out to be one of the most compelling features of this method.

The withdrawal and retention thresholds bracket a zone in which additional data, such as the existence of digitized copies and where they are held, are collected to make a decision. This category was established initially because the thresholds do not perfectly align, and we
had intended for items in this middle group to be candidates for referral to a decision-maker outside of preservation. This category drew on our assumption that this group of materials would benefit from further knowledge, such as collection priorities or local user needs, in the preservation decision-making process. In practice, only about 1% of items fell into the middle category and almost none had their default decision changed by a secondary reviewer. The value of this group warrants further consideration for future implementations of this decision-making framework.

**Formulas for Replacement and Withdrawal**

The moderate formula indicates replacement of items with fewer than twelve (12) global holdings, fewer than three (3) California holdings, and zero (0) additional UC holdings; withdrawal of items with more than twenty-six (26) global holdings, more than five (5) California holdings, and more than two (2) UC holdings; and to ask collection managers to review all other items. Based on Yano’s model, and assuming an annual loss rate of 1% or less, these thresholds allow us to estimate the probability that, after taking the local preservation action, a complete copy can be found or reconstructed in 2100 CE. For California alone this probability comes to approximately 78%, and surpasses 99.5% across global holdings.

The aggressive formula replaces items with fewer than four (4) global holdings, fewer than one (1) California holding, and zero (0) UC holdings, and withdraws items with more than twelve (12) global holdings, more than three (3) California holdings, and more than one (1) UC holding; all other items will be reviewed. This formula reduces California’s perfect copy survival probability to about 65% and places the world-wide survival odds at 92.5%
Application of the conservative formula leads to replacing items with fewer than sixteen (16) global holdings, fewer than four (4) California holdings, and one (1) other UC holding, and to withdrawal of items with more than thirty (30) global holdings, more than ten (10) California holdings, and more than three (3) UC holdings; all other items will be reviewed. This formula produces a California state-level perfect copy survival probability of 87%, based on four (4) copies, and world-wide survival odds upwards of 99.9%. Table 1 shows the withdrawal and replacement thresholds for each formula.

[Insert Table 1 here]

Table 2 shows the WorldCat holdings and associated survival probabilities for each scenario described above and listed in Table 1. In Table 2, the survival probabilities can be read as risk tolerances. They could also be read as a statement about the level of risk at which we feel compelled to hold the line against further loss (replacement threshold) and above which we feel confident in the system’s ability to absorb the loss of UCLA’s damaged copy – and still maintain access well into the future (withdrawal threshold). This implicit reading of the moderate threshold, which UCLA has been developing for actual application rather than theoretical study, suggests that “UCLA Library evaluates preservation risks out to the end of the century, always takes preservation action when a work’s chance of survival drops below 90%, and never withdraws materials unless there is a 99% or better chance that the global library system can preserve that work until the year 2100 CE.”

[Insert Table 2 here]
Table 3 shows the range of probabilities for the library network to be able to produce a single perfect copy in 2100 CE, given an initial number of starting copies, based on Yano’s study. The table shows probabilities for two loss rates. We have strong reasons to include the 1% loss rate projection as our default assumption for planning purposes. This level closely matches our internal estimates of loss based on interlibrary loan data, circulation data from the Southern and Northern Regional Library Facilities, and data from the UCLA Library catalog and circulation records. A second section of the table shows the equivalent risks at a 2% loss rate, however, to provide additional context and a worse (if not worst) case projection. In the interest of being risk-averse, we employ the 2% loss scenario in Table 2 (above) and in other analyses throughout this report.

[Insert Table 3 here]

Accurate assessment of risk is most important at the lower threshold, where the decision branches between defaulting to replace the item in question, or to send the item for additional review. At the withdrawal threshold, the difference in survival probabilities narrows to about 1%. Even that 1% difference causes little real impact on survival probabilities in the sample we studied, because very few items actually comprise holdings precisely matching the threshold level. Table 4 shows the median, average and maximum holdings levels for global, California, and the UC system across the entire sample. At the median global holdings level of 61 copies, the difference in survival probability between the 1% and 2% loss rate projections shrinks to .002%. The high level of median holdings
supplies important information for our risk assessment. Even if the real rate of loss is twice our estimated rate of loss, 2% instead of 1%, at least half of the materials considered in preservation review will maintain the same survival outlook.

[Insert Table 4 here]

Figure 1 shows the relationship of the 1% and 2% loss rate projections to one another. Both projections approach 100% probability for the survival of a perfect copy in 2100 CE at the 26 copy threshold for withdrawal decisions that we use in our moderate scenario.

[Insert Figure 1 here]

When we applied the three different formulae to our study materials, the findings emerged as expected; the aggressive formula led to more items withdrawn, fewer items replaced, and slightly more items reviewed. The conservative formula led to more items replaced, fewer withdrawn, and more reviewed. Table 5 shows the numbers of items, in a total sample of 1,408 items, that receive default decisions of withdraw, replace, or review.

[Insert Table 5 here]

**Implications for Collection Strategy**

Although the data collected primarily facilitate expedited decision-making and support preservation management, they can be analyzed with a view to collections management and their broader implications for the library’s mission of stewardship. For example, a telling indicator within the preservation review materials reveals that median
holdings are lower than average holdings by a factor of roughly three (Table 6). This suggests that a relatively small number of very commonly held works inflates the average. Figure 2 illustrates this pattern, commonly called a “long tail” after the book and article by Chris Anderson.36 Significantly, Figure 2 is truncated at 400 holdings, the 90th percentile for holdings. The top 10% of global holdings for these materials cover an order of magnitude change from 407 to 4,085 holdings.

[Insert Table 6 here]

If we examine the holdings for the entire set of 1,408 items, a clear long tail trend appears, with the exception of the UC holdings. The fairly flat curve of UC holdings likely arises from the small number of UC campuses compared to the number of libraries in the world. This also fits with the shift in UC policy towards limiting the purchase and maintenance of multiple copies across the system wherever possible. Given the origin of this dataset in preservation review, especially, the low UC-wide holdings remain consistent with expectations and UC collection development policies.37

[Insert Figure 2 here]

To view these data in another way, we plot global holdings against the lower of UC or California state holdings. In Figure 3, global holdings are plotted logarithmically, since they greatly outpace state- or system-level holdings, and the resulting scatter plot is divided
into quadrants that show the default actions for each combination of local or world-wide scarcity.

[Insert Figure 3 here]

Materials that are widely held from a global perspective but scarcely held within the local system become particularly visible in this presentation, where they cluster in the upper left hand quadrant. As noted above, adding replacement thresholds for local or consortial groups makes almost no impact on system-wide risk to the loss of print materials. However, it transforms preservation decision-making and the ability of a given library or system to independently meet its users’ needs.

Forty percent (40%) of the materials in our study falls into this locally scarce but globally common set of materials. This figure poses an important counter-balance to the idea of a single, centrally planned, national collection of record. Our analysis suggests that preservation strategies may be best articulated on a regional or consortial basis, with national goals emerging from those efforts. To illustrate how local collection management decisions and efforts to improve the effectiveness of a single preservation operation can create positive global outcomes, we turn to two scenarios that can be played out using the data and risk projections presented thus far, by adding additional retain and withdraw thresholds for consortia or regional collections.

The Influence of Regional and Consortial Holdings

In the first scenario, we add thresholds for the number of copies in the UC system and in the state of California. The differences between the withdraw, replace, and review
thresholds in each of these scenarios are small, as Table 7 demonstrates. The \textit{Difference} column of Table 7 shows how much closer the UC and California numbers are to each other, as compared to the global numbers, effectively narrowing the band of “review” materials needing input from outside the preservation department. Consequently, fewer items require secondary review in scenarios that include UC or California holdings, even though the system-wide risk of loss (driven by the global “withdraw above” threshold) stays the same.

[Insert Table 7 here]

The total numbers of withdraw and replace decisions do not vary greatly between any of the original formulas; however, when one of the holdings groups is removed from the formula, the differences increase. Table 8 shows the percentage of each type of decision in the three formulas with and without UC holdings data.

[Insert Table 8 here]

If UC holdings are ignored, but other constraints remain, the difference between the categories shifts by 16% towards withdrawal of material. Purely within the boundaries of this model, and if it had no obligations to the UC system, UCLA could, in effect, justify an 11% increase in the rate at which it deferred preservation responsibility to other institutions. This is not, in fact, the aspiration, policy, mission, or vision of the UCLA Library. And fortunately, we also have some evidence that adding consortium or regional thresholds creates important benefits for the local library: labor-savings, by reducing secondary review of materials, and an additional level of protection against loss of materials. Taken together, these indicators
provide a useful object lesson in how to situate evidence-driven practice in library management. These data allow us to explore many possible paths and outcomes, but as a matter of institutional mission, UCLA Library intends to serve as a collection of record that “develops, organizes, and preserves collections for optimal use.”

Read from an intention to preserve, these models reinforce the importance of explicitly declared networks for preservation and provide an incentive to associate ourselves with other collecting institutions, the better to safeguard against withdrawal activity that would threaten our mission. In effect, the constraints of system-wide policy and obligations to local or regional groups enjoin us to a greater degree of collective preservation effort.

The inclusion of California holdings information makes less of an impact than the addition of UC data, but it keeps items out of the review category in the moderate formula, and leads to a slight shift in the proportion of materials retained or withdrawn. The review category in the moderate formula increases significantly if the California holdings data are ignored, illustrating a common theme that having more decision-making data about each title allows for more default actions to withdraw or retain, and fewer review decisions, shown in Table 9. The effect of removing the California holdings, but leaving UC and global holdings, is less dramatic than removing the UC holdings.

One conclusion is that California is simply a large enough state with enough major libraries that it operates as a good proxy for the world. This may be less true as the scale of analysis shifts from this case study to the entire collective-collection, but the alignment between these data and the work on mega-regions merits further attention. States, regions, or other large affinity groups may provide a valuable scoping mechanism for planning preservation networks. Such groupings may be large enough to approximate the global
patterns of retention necessary for the long-term survival of materials, while also being small enough and having enough shared interests among stakeholders to organize collaboration. Indeed, at this scale, there are likely to be pre-existing collaborative ventures that can be leveraged for print archive efforts.

[Insert Table 9 here]

Table 10 shows how removing either the global holdings or the local holdings (UC and CA) from the moderate replacement thresholds (replace below 12, withdraw above 26) leads to a more conservative outcome than a scenario that only looks at global holdings through WorldCat on the one hand, and a more aggressive approach than an exclusively local scenario, on the other. Considering only global holdings leads to 30% more withdrawal but 50% less automatic replacement. Considering only local holdings, in this case California state holdings and UC system holdings, leads to slightly less (-19%) withdrawal and almost identical replacement outcomes (-2%).

The elimination of either the global or regional holdings increases review cases as well. Although the percentage change is staggering (782-1100%), the actual number of review cases only becomes significant where global holdings are not evaluated. In that instance, there would be 150 items for review, 11% of the total sample. As we expanded this work from a pilot project with 376 samples to this case study of 1,408, we consistently found that the review category diminished in importance as we added groups of holdings data. From a purely practical view, within this model of preservation review, one could argue that
it is worthwhile to always evaluate the global holdings just to reduce the amount of selector review required, and by happy coincidence, this also helps to avoid global scarcity problems.

[Insert Table 10 here]

The impact of these scenarios on the review category has been of particular interest for two main reasons. First, keeping items out of the review category results in fewer backlogs in the preservation department and faster progress of materials through the preservation decision-making process. Second, placing items in the review category shifts some decision-making work from preservation staff to collection managers. The time that it takes preservation to get the UC holdings from WorldCat is small and easily measured, unlike the time it takes for a collection manager to receive and review an item in order to make a decision based on their knowledge of or beliefs about its usefulness to a field of study and relevance to the collection.

Another factor we considered in decision-making is whether or not an item appears in HathiTrust. We originally planned to use HathiTrust data to resolve items left in a review status for several months, so that if collection managers had not made a decision on review materials and a full-view HathiTrust copy were available, we would revert to a withdrawal decision. Although the small number of materials requiring outside review in our study has made this a moot point, the rationale behind this consideration serves an important purpose. Because we evaluate reformatting options at this stage in the preservation review process, and since these items already fall in a lower-risk category (above the immediate action threshold, but below the automatic withdrawal threshold), withdrawal of the damaged
copy in-hand does not forego the system-wide possibility of creating a better digital version in the future.

The results of fully incorporating HathiTrust data into our decision-making are shown in Table 11. Using the moderate formula as our starting point, we subsequently used the presence of a HathiTrust copy to change review decisions into withdrawals. In the scenario that only uses global holdings, changing our decisions based on the existence of a copy in HathiTrust only affects outcomes for four volumes. However, in the localized scenario that includes the UC and California holdings, this approach greatly reduces the number of volumes left in review status, from 150 to 9, a 94% decrease. This application of HathiTrust data suggests some potential for consortial networks to do effective internal coordination of print and digital preservation efforts.

[Insert Table 11 here]

Although this model of preservation decision-making shows many advantages, we anticipate concerns for a hypothetical situation where the library withdraws the damaged copy in its collection, but no other libraries with holdings of that title have a policy that would ensure they retain their copy. In the moderate formula, UCLA withdraws when there are more than 25 extant copies, so the specific version of this concern would be that two dozen other libraries will lose their copy or withdraw it without any consideration of the network impact of that action. The likelihood of that is very difficult to determine. Although we might assume it to be quite low, we have no definitive data to affirm or negate this assumption. This dilemma supports an additional, compelling justification for having a
second set of replacement criteria built around a region or consortium, in this case California and the University of California system.

As we discussed above, adding the consortial checks reduced the quantity of materials withdrawn as a default by 30% in our study (see Table 10). Further, as Figures 1 and 2 show, a substantial amount of material that could be withdrawn without affecting system-wide survival rates must be retained in order to protect consortial preservation and access. This point bears some additional emphasis, given the amount of activity focused on network-level collection development. For several of these network efforts, the wish to avoid duplicate maintenance costs and the possibility of recovering space in library facilities form the chief motivations; these projects focus primarily on journal back files.\textsuperscript{40} Robert Darnton recently made a broader argument for a renewed national library building effort; recent announcements by the HathiTrust and Internet Archive indicate their engagement with the idea of building a print archive to mirror their digital holdings.\textsuperscript{41} None of these projects specifically say that one and only one copy of a work is the end game. However, most of these efforts center their activity on a small number of copies, identified from a network level, and placed into trusted archives on behalf of the community. We endorse the core concerns identified by these projects’ leaders, but our analysis suggests that at least a dozen regional or consortial efforts, rather than a few national efforts, will lead to more favorable outcomes. We believe that localized efforts are more likely to see implementation and will be easier to maintain, since they build on existing communities and established operations. Our data show that multiple systems acting independently can provide reliable preservation, especially if they have a means of disclosing their actions.\textsuperscript{42}
As we test different replacement scenarios, we cannot simply develop an internal matrix for decisions and declare success. Preservation operations have two stewardship responsibilities. One is the basic fiduciary responsibility to expend funds in an effective and transparent fashion. The other is the larger goal of ensuring that those funds actually support the survival of cultural and artifactual heritage. The methodology described above places both stewardship responsibilities into a measurable framework that supports meaningful comparison and collaboration between institutions.

Further Research

Our approach to making preservation review decisions only governs UCLA Library, not the decisions of the other holding libraries in WorldCat. As discussed above, we selected replacement and withdrawal thresholds to allow for a margin of safety against the possibility of system-wide loss of access to a work. We further hedge against the risk of loss by applying consortia holdings as a decision-making element, so that even if a work is widely held, we maintain a copy in the state or UC system. A secondary benefit arises as well, since the more holdings groups we include in the formula, the fewer items fall into the review category. We have not evaluated the effect of adding additional real or hypothetical groups to this analysis, but that points to a promising direction for future research.

Further studies should be conducted to determine which pieces of data are the most important to collect and what replacement thresholds most closely mirror collection managers’ levels of risk tolerance. By providing proposed decisions to collection managers without identifying whether the aggressive, moderate, or conservative protocol was used, it would be possible to track the way in which their decisions align with the pattern of default preservation decisions and to see which data most strongly correlate to their decisions.
Additionally, areas where managers consistently decide against the numerical indicators offer valuable clues about the library’s collecting interests and intentions. This information may be key to developing more effective assessment and decision-making routines. Ideally, we will collect only data with the strongest relationship to a desired replacement pattern. For example, if we determined that decisions based on the UC system holdings data correlated with 95% of the collection manager decisions, and the global holdings data correlated with 94% or 96% of the collection manager’s decisions, we might question the utility of collecting global holdings data at all. As the data-driven decisions start to more closely mimic collection managers’ decision-making, we may also re-evaluate the utility of the review category between the withdrawal and retention thresholds. Eliminating the review category would further the goals of reducing a backlog in preservation and of making efficient, timely, evidence-based decisions.

In addition, better analysis tools for the WorldCat database would facilitate the investigation of questions such as overlap of holdings and classes of libraries that hold various types of materials. For instance, if we found that of the 32 copies held for a given title, the 31 non-UCLA copies were in small school and public libraries, we might conclude that the preservation responsibility were ours. Contemporary library collection management strategies increasingly incorporate cross-collection availability patterns and the potential for systematic collaborations. For this, the holdings data in WorldCat needs to be much more accessible. The WorldCat Collection Analysis tool made some progress in this direction, but its restricted access and high pricing impeded its use; its publicly available documentation and interfaces did not disclose clearly how one could create selected title lists for comparison, rather than entire library collections; and the tool provided limited ability to create
customized datasets that would be useful for analysis in external tools. Collection Evaluation replaced the Collection Analysis tool in late 2013, after this study. We have not yet evaluated this new tool for our purposes.43

Conclusion

Library preservation activities can benefit from new techniques for management and new technologies for treatment, yet the mission of preservation remains inherently cautious and somewhat conservative. A book kept is not a book lost. Preservation administrators must be keenly aware of the opportunity costs involved in delayed or slow-moving decisions. Still, they must pay equal attention to the opportunity costs involved in working only in data-driven abstractions and neglecting the individual value of the items in their care. UCLA Library, for instance, encompasses a collection of upwards of 9 million volumes, and its archives and manuscript collections measure in linear miles. Against this backdrop, the effort described above, which evaluated about .02% of the collection, denotes little in a purely numeric sense. This project’s primary value stems from its addressing the review needs that emerge from the subset of the collection used in day-to-day library business. It prevents the accumulation of large amounts of unusable materials with no clear outcomes assigned to them, while freeing limited local resources to be spent in a way that provides the greatest return. By using a risk-assessment model to guide our decisions, we can ensure that we spend our monies to immediate, positive effect and avoid costs that can be safely pushed elsewhere in the global library network and to a later point in time.

The broader secondary value of this project derives from its ability to both improve network-level library services and to provide models for local action within large-scale collections strategy. In the work to date, we have identified several hundred volumes that
would otherwise be excluded from large-scale digitization efforts. We have identified several volumes of substantial scholarly and artifactual value appearing unique or so scarcely held that UCLA Library became the only realistic point of access in North America. This has been achieved through a balancing of efforts. We hedge one method—holdings-based analysis of an appropriate scale to the global library enterprise—with others—item-level review before withdrawal and targeted external review by specialist librarians—so that objects with scholarly or artifactual value do not get overlooked. This combination of a broad view of collection management coupled to a clear set of local needs is highly effective in preservation management and suggests a strategy for other library collection management efforts.
References and Notes

Note: All electronic available through open http were accessed and archived as of 16 August 2011. See http://www.jacobnadal.com/162 for information about archived content.


7 Ibid.


9 Ibid., 115.


19 Ibid., 6-7.

20 Ibid., 13.


22 Ibid., 1281.

23 These inferences are drawn from the published literature on library condition surveys as well as a collection condition survey project conducted by the UCLA Library Preservation Department. More information on this effort is available at [www.jacobnadal.com/82](http://www.jacobnadal.com/82). For other library condition surveys, see Thomas Teper, “Building Preservation: The University of Illinois at Urbana-Champaign’s Stacks Assessment,” *College and Research Libraries*, 64 no. 3 (2003): 211-227, and Gay Walker, Jane Greenfield, John Fox, and Jeffrey S. Simonoff, “The Yale Survey: A Large-Scale Study of Book Deterioration in the Yale University Library,” *College and Research Libraries* 46 (March 1985), 111-132.

24 These calculations assume an average book thickness of 1.5", with stacks assembled from six 36" shelves, filled to 90% capacity. These dimensions are drawn from UCLA Library’s internal collection surveys.


27 The adhesive used in Post-it notes is described in Silver, Spencer Ferguson, “Acrylate Copolymer Microspheres,” US Patent 36911140. This patent application describes a method...

A photo gallery of example items from preservation review is available at: http://www.jacobnadal.com/###

UCLA buys replacement copies listed by their booksellers as “Very Good” or better and with no defects noted. To date, we have used every one of the more than 100 books purchased, with only two requiring minor stabilization treatments. See IOBA Book Condition Definitions for a listing and discussion of bookseller’s terminology: Independent Online Booksellers Association, IOBA Book Condition Definitions, accessed August 16, 2011, www.ioba.org/desc.html.

One of the critical issues latent in preservation decision-making is the research library’s role in the maintenance of material culture, outside of its dedicated special collections. We have two working assumptions in this area. One is that within the library staff, the preservation and conservation staff is the best-equipped to evaluate the material significance of the library’s artifacts. The other is that the severely damaged materials in preservation review are of little or no value to the study of the material culture of the book. Of course, severely damaged books are of value as a study collection for preservation and conservation, arguably a related field to material culture studies, but in this regard their presence in preservation review means they are accessible to the scholarly community that needs to interrogate them.

Yano, 11.

Because of restrictions on providing full-view access to in-copyright materials in the HathiTrust, the copy in preservation review is sometimes placed in an enclosure and transferred to the SRLF, in lieu of withdrawal, until the emergence of a consistent means of managing full-view access for in-copyright materials in HathiTrust.

Yano, 33-34

A perfect copy in this instance retains all of its contents so that it could be used to create a usable digital version. Perfect copies are assumed to be subject to the normal decay that affects all cellulosic materials, a problem preservation efforts guard against through preventative treatment such as deacidification and environmental controls.

These data were collected by the Persistence Policy Implementation Task Force, on which Mr. Nadal served as a member, in the course of developing replacement guidelines to support the UC Libraries Persistence Policy. The Persistence Policy and related reports are available at http://libraries.universityofcalifornia.edu/planning/


42 It is worth noting that what we propose is, in effect, the same model as the LOCKSS project, only for print rather than digital resources. It is no mere coincidence that LOCKSS networks become robust when there are 10 or more LOCKSS appliances in the network. More information on LOCKSS is available at: http://lockss.stanford.edu.