AC 2012-4644: WEEDING THE COLLECTION: AN ANALYSIS OF MOTIVATIONS

Peter Zuber, Brigham Young University

Peter Zuber is the Engineering Librarian at Brigham Young University. Apart from liaison and collection responsibilities, he has evaluated and implemented virtual reference services and helped in the redesign of the library’s website as well as the integration of a new federated search engine. As Chair of the library’s Research Behavior Group and User Studies and Assessment Team, he worked with librarians to discover user preferences and habits and how they impact current library services and tools. Published papers include topics such as search engine constraints, open access, institutional repositories, and remote services. Before joining the university, he was Director of Imaging and Process Engineering for T/R Systems in Atlanta. He holds 29 patents in the areas of electrophotographic engine design, color theory, and digital image processing. He received his M.L.I.S. from the University of Southern Mississippi and his master’s in imaging science and engineering from the Rochester Institute of Technology in New York.
Weeding the Collection: An Analysis of Motivations, Methods and Metrics

Abstract

“The traditional (or Alexandrian) model of the library is based upon the mistaken intuition that to be good a library must be vast and always growing.”¹ Weeding of library collections, whether printed or electronic, whether by choice or compulsory, creates opportunities for criticism and growth. Where the literature suggests that starting a weeding project is close to last on a librarian’s list, events beyond a librarian’s control can force a weeding action. Space reallocations, time constraints, or e-provider policies imposes risks that threaten the attempt to maintain a viable collection. Sometimes the choice is voluntary; but if not explicitly defined as part of a formal collection policy, the rationale used can vary, and if the librarian is inexperienced, the methodology risks being incomplete or at worst, inaccurate. Although varied criteria exist to help determine what to weed, successful results depend on proper application. Considerations such as local historical interest or the ubiquitous nature of the content will impact decisions as much as typical variables such as physical condition, usage, or curricular integration. Additionally, within the disciplines that many engineering librarians support, the volatility of those subject disciplines can vary widely.

Paying particular attention to engineering-related subject disciplines, this paper will outline historical motivations that drive weeding initiatives, describe varying methods previously employed, and propose if a systematic rationale for applying metrics to determine candidate materials as a function of discipline could be developed and successfully employed. By listing potential variables, suggesting which ones would be emphasized for particular collections, and contrasting their application in a manual versus an algorithmically applied routine, the paper will propose a best practice methodology.

Introduction

For the new professional, the practice of weeding a collection can appear daunting, especially when experienced colleagues express varying opinions on efficacy or desire. Step-by-step methodologies are rarely found in the literature, leaving the professional to consider descriptive terms such as “cluttered,” or “unattractive,” or “outdated materials” without much guidance. The need for weeding is almost always encouraged, but buffered with warnings of difficult decisions, risk, and political incorrectness.² For an engineering professional accustomed to industry standards with detailed procedures, quantifiable outcomes, and descriptions on how results were obtained, the ACRL standards for college libraries was less than encouraging with its initial singular statement: “collection currency and vitality should be maintained through judicious weeding.”³ Some criteria were offered, including discarding materials “which have outlived their usefulness” and those for which “a clear purpose is not evident.” Where this same standard can be found replicated with some variation within various college and university collection development policy statements, it is challenging to find concise and universal agreement on what “usefulness” or “clear purpose” really mean. The lack of rigid guidelines of how to weed might be considered current practice given the set of conflicts, challenges, and varied motivations weeding presents. Historically, many of these motivations have been
discussed. Understanding these sets of motivations is of value if a more structured approach is to be proposed.

Collections should be kept intact

The Alexandrian model mentioned previously equates a fear of loss (notably and not without reflection), at the hands of barbarians. The risk of loss and the perceived difficulty in recovering materials is not trivial and likely plays a major part in the perception of some that all collections should be kept in perpetuity.\(^4\) In addition, the fear of program or curriculum changes in the future may tend to temper ideas of ridding the collection of little-used titles that may reveal their value in the future.

Budget / Comparisons

Typically, the size of the collection is often used as a metric when considering budget allocations or making comparisons to other institutions. Additionally, the collection size may often be “considered a criterion of the quality of a library.”\(^5\) If so, this suggests a strong, defensible rationale would be needed, especially where funding or rankings are concerned, if titles were to be removed. Consequently, strong motivation exists to maintain the status quo or even increase holdings rather than weed and have to assume a defensive posture.

Policy

Some libraries provide little or no information in their collection policies on weeding, procedural or otherwise. This lack of direction may be due to an inability to fully articulate how weeding should be done for each collection area, or reveal an emphasis on acquiring rather than deleting content. Some policies do state that weeding should be done; some even provide percentages per annum expected. Going a step further, some policies state percentages in terms of anticipated future use, requiring the librarian to forecast usage. While this may not be entirely impossible, it does create the need to study usage patterns, perform sophisticated statistical analysis, and may include crossing fingers. After these calculations, questions may still remain, not only on what is wanted, but on how librarians can get answers and what tools they can use.

Politics

Seeing a library throwing out materials appears at first glance to be highly inappropriate. Manley states that the practice is perceived as politically incorrect, and that citizens “go nuts when they find out that libraries discard thousands of books every year.”\(^6\) With more electronic access, greener initiatives, duplication, and OCLC interlibrary loan support, this issue becomes less volatile.\(^7\) Alternatives also exist in charitable organizations or recycling. However, some patrons will always prefer a book; greener alternatives require more planning; and time is often at a premium given other tasks at hand.

Librarian’s time or motivation
Studies have reported the tendency of librarians to find little time for weeding or to maintain the motivation to work on weeding. Wallace states that the use of conventional weeding techniques, meaning judgment based decision making, can require several minutes per title. As the number of volumes increases, the dedicated time required for weeding can reach into the thousands of hours, perhaps even years of professional level effort. This work pressure, along with the other tasks a librarian must perform, can place weeding very low on the list of things to do.

Motivations for weeding are numerous. Although some categories of materials require careful analysis to arrive at a decision, others are self-evident. Within the latter category are holdings which become redundant due to electronic acquisitions such as journal subscriptions, printed abstracts, or books which have their equivalent in electronic form. These are likely candidates for removal. Among the category requiring careful analysis are unique holdings that are in or contain obsolete media such as VHS, floppy disks, or cassettes. These materials are unique in their dependence on older technologies and their likely inability to migrate to newer formats due to copyright. These may require commitments in sustaining equipment for their use. Other titles may contain both print and not so obsolete media such as CD’s or DVD’s. Programming books are typical. These titles may have been cataloged separately for convenience or protection. If consolidating these items is required, often the only recourse is to do so and risk the consequence of damaged media from improper handling. Other alternatives include removing them or sending them to auxiliary storage. Finally, there is the remainder of the collection. For these titles, three motivations for weeding follow: space, increased circulation, and ease of use/patron perception. These three serve as an encapsulation of several motivations suggested in the literature. For the most part, other motivations are variants of these major categories.

Space

In the pre-electronic era, space concerns were paramount. Shelving space was at a premium to the point that some crisis-based rationale included weeding titles that were “thicker” or whose volume set size took up large amounts of shelf space. In most situations, electronic acquisitions and auxiliary storage schemes have provided enough shelf space relief that removal for size rather than content is less of an issue. This is not to say that space is still not a primary concern. Some might argue that the ability to load thousands of books on virtual “shelves” might be motivation not to weed the physical collection. But for a contemporary library, the “library as a place” is an important element in its service model. Group study areas, reading rooms, computer workstations, media centers, “no shhh” zones, and cafes are enticements for patrons to come into the building. These areas require space. Difficult to justify is occupying space with materials that experience no use when the space can be better used for the patrons’ benefit. Additionally, and related to the patron perception addressed below, the more concise and efficient the physical collection becomes, the more likely the patron will return to use it.

Increased Circulation

Librarians recognize and the literature supports the conclusion that effective weeding increases circulation. McKee demonstrated strong relationships between “declining book stock and increasing circulation,” and Poller has shown that weeding efforts in specific areas of a
library collection were directly correlated with circulation increases in those same collection areas.¹³

Ease of Use / Patron Perception

Electronic access provides rapid and convenient searching, virtual browsing of physical shelves, and full text content of select resources. Consequently, the ease of use motivation might be considered less impactful when compared to the practice of years ago; card catalogs and browsing physical shelves. But removing outdated, irrelevant, or less impactful materials, whether physically or electronically, can still significantly improve ease of use. Mooers’ Law remains a strong directive:

“An information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it.”¹⁴

Apart from writing an improved relevancy algorithm for a library’s search engine or influencing cataloging schema, the best practice for a librarian is to provide content that is current, scholarly, relevant, integral and compelling to the patron. By continual application of this practice, a self-perpetuating cycle toward greater patron satisfaction will develop. For example, titles that provide ideal content will naturally be those that experience consistent usage and those considered untouchable such as classic texts, thesis/dissertations, local authors and so on. Then, for a given collection area, as only those titles are kept, the catalog is weeded of irrelevant, useless titles. Searches will naturally tend towards higher relevancy. Thus, ease of use is improved and the collection will trend toward a higher usage percentage. The higher the usage percentage, the more useful the collection will appear, and the patron perception of the library as an important information center will increase. The “patron satisfaction cycle” may initially require a significant removal of titles but, through consistent weeding adjustments over subsequent years should perpetuate and provide consistent results.

Methods

Two primary methodologies exist when considering how to weed a collection, one qualitative and the other quantitative. The qualitative method is largely founded in judgment-based decisions of the librarian. These include the librarian’s experience, interpretations of policy, curriculum, cost, perceptions of needs, value, future use, and so on. The list of variables to consider is extensive and can vary depending on implementation. Slote reports that this method most likely represents the “majority opinion in the country.”¹⁵ It is easy to see why such an approach would have strong adherence. Given the motivations listed earlier, as well as anticipated repercussions if weeding practices backfire, the experienced professional would be the likely choice to make these decisions and address such concerns. However, with an exclusively subjective-based system, outcomes would likely vary with the decision maker, especially if guidelines or directives are vague. Each of the previously listed motivations could be interpreted differently based on the perceptions of the administrator, the librarian, or the public. For the administrator or librarian, the exclusive use of judgment-based methodologies for weeding a collection creates opportunities to “change the game plan” when contingencies arise. At its worst, the Alexandrian example suggests that destroying (weeding) a collection is highly
undesirable. Colleagues may see it as inappropriate within a scholarly environment. Political pressures, time constraints, and lack of clear justification can create emotional components that alter outcomes in an attempt to protect against perceived risks.

In addition, a propensity for caution when weeding could be reasonably attributed to the challenges inherent in determining how to effectively weed, especially if judgment-based criteria is the sole determinant. In a study of Purdue University libraries, Lister suggests that “Intellectual weeding policies, which require judgment and are based upon somewhat intangible variables, usually turn out to be time consuming, require judgment and qualitative attempts to predict future use.” This serves to introduce a dilemma: the more an approach depends solely on qualitative judgments the more additional variables tend to be introduced. Segal describes this dilemma of increased accountability: “this burden forced librarians to ask just one more question and consider just one more factor before cancelling a subscription.” Using a system that leads to increasing variability will create indecision and frustration that over time becomes incapacitating. This self-perpetuating mechanism can become a perfect storm of conflicts and inaction. Fohl reported that it took six months for just the weeding portion of their collection management effort due to faculty indecision, librarian judgment, bargaining, pleading, and so on. The same judgment-based, qualitative decision making approach rarely produces the same incapacitating effect when content is being acquired as compared to when content is being weeded. The emotional component present during the removal of titles may be absent in their acquisition.

The second methodology, quantitative, is based on objective measures which fail, meet, or pass predetermined criteria. Circulation rates, in-library use, age, or search statistics are the commonly used variables. This system is ideally suited for computer application, and if the algorithm is deemed acceptable, can rapidly process and produce results within minutes. Historically, objective-based weeding algorithms have been created that focus either on search statistics, citation frequency, or usage rates. Of these, the most accepted and widely used variable is use. Slote is well known for “shelf-time periods,” or the time a resource experiences between uses over a period of time. Simply put, titles that had a small number of uses over the last “n” years would be candidates for weeding. The cutoff percentage and time periods used are variables that can be determined using shelf-time calculations for each collection area. Slote found that the age of a title was a low predictor of future use, but shelf-time period was a very strong predictor. Other studies have used Slote’s method and found similar results. Usage rate determination is perhaps the best indicator of patron preferences within the collection. Whether these include circulation statistics or in-library use, both clearly show what a patron wants and where the collection is experiencing its highest return for the investment.

Trueswell offers an idealized way of leveraging usage rates and demonstrating how it can have a significant impact on weeding collections and setting collection size. Trueswell assumed from previous work that only a small portion of an entire collection was frequently used, and this small portion would satisfy the needs of most, if not nearly all, users. This small portion of the total collection was defined as the core collection. For example, if a usage study found that 99% of patrons frequently checked out titles that amounted to about 30% of the total collection, this 30% would be considered the core collection. Removing the remaining 70% of the total collection would drastically reduce the collection size, yet still satisfy about 99% of the
circulation requirements. Although a simple calculation, the implications are profound. Studies at a technical library revealed that 99% of the current circulation came from a population of titles totaling 25% of the collection, used within the last 8 years. To illustrate discipline differences, another library realized a 20% core collection but it was based on an 18 year window. Given some discretion in protecting certain non-eligible titles, this approach demonstrates the power inherent in usage as a variable.

Other variables are available for use in a quantitative methodology, but the question to consider is whether they provide reliable correlations to weeding objectives. For example, the absolute age of a title was shown in previous studies to have poor correlations to future use, but studies have also shown that as newly acquired titles age, their first time use is reduced. Kent found that as new titles get older, first time use reduces significantly each successive year.25 For example, during the first year only 26% of newly purchased titles were first used. The second year saw 17%, the third 6%, and so on. In total, 40% of new titles were never circulated over seven years. Given this type of circulation performance, Trueswell suggested that librarians should “remove all books that have not circulated during the previous eight year period.” This suggests that a “last use” metric for all titles and a “new title age” metric for new titles might be viable.

Apart from methodologies, other considerations such as discipline differences should be considered. Slote suggests that collection size in an area of study should match the circulation patterns in that same area in order to optimize investment.26 In other words, using Trueswell’s terms, the total collection size should be closely matched to the core collection size. Additionally, there are volatility considerations as a function of discipline or subject matter. Today, the computer science collection in general is more volatile than mechanical engineering, while the alternative energy topics show increased usage in electrical and mechanical engineering. Shelf-time period calculations can be applied for discrete collection areas, providing discipline dependent percentages.

Looking at Engineering disciplines in general, Lucker and others found that “the predominant interest of the majority of users, especially in science and engineering, is in the most recently published information, with periodical literature dominating. There are, however, substantial numbers of students and faculty interested in the historical aspects of a number of fields, particularly technology, architecture, and science.”27 For each of the objective variables discussed such as usage, age, last use, first use, there should be a consideration of discipline, and perhaps in some cases, consideration of topic. Call number ranges are recommended to create discrete “total collections” that exhibit similar volatility in use and currency. For example, QA76 + TK 5105 may be the “total collection” for computer/website/database titles for purposes of weeding.

Given these approaches, how could using either or both methodologies leverage the important aspects of a librarian’s judgment, along with the interest of varied patrons? Mosher, in discussing the management of library collections, stresses “the importance of using more than one criterion or method in identifying material for deselection and advises that a mechanical technique such as use studies or citation frequency be used in addition to the judgmental considerations of program needs.”28 Given the need for consideration of both methodologies, and
the inherent strengths and weaknesses of each method, this paper discusses methodologies in terms of patron needs and wants.

Patron “needs,” for the purposes of discussion, simply represents the position that weeding should be motivated by what the librarian perceives the patron should have. The librarian makes a determination of what the collection should contain based on this criterion. However, the patron “needs” approach is predominately based in qualitative methodology. Consider some of the variables used: perceptions of future use, perceived responsibility to the user, curriculum needs, space, costs, collection policy, accreditation requirements, electronic preference, duplication, service, organization, accessibility, and scholarly content. Since these criteria require the judgment-based experience of the librarian, best practice would be to apply qualitative judgments at the front end, in other words, at acquisition, especially given the emotional conflicts qualitative judgments produce at weeding. In fact, one might argue that the selector is essentially making a “weeding” type decision every time a title is not chosen for purchase. Indeed, Segal mentions that “deselection implies undoing a past action – one of selecting.” In this case, one might better relate acquisition to weeding by considering it the “de-selection” or “un-selection” or “non-selection” event. As such, qualitative methodologies should be applied primarily at acquisition as the selector can make a more reasoned decision at this time. This is not to suggest that qualitative judgments are excluded from the weeding process, but their role should be minimized. In addition, selectors will find they will be better informed of future usage when they know the trends of particular collections from the usage studies that will be performed on the existing collection from weeding reports.

Patron “wants” is assessed by circulation statistics, by survey, or by assessing search results. Since usage statistics and similar data are employed, the “wants” model clearly uses quantitative methodologies. Many of these data can only be evaluated by assessing an existing collection. Since quantitative methods are easily applied to existing collections, quantitative methodologies should play an important role in weeding existing resources. Weeding with a wants/objective-based model leverages gathered information about how resources are being used and provides insights on collections to a highly discrete level. Quantitative methodologies also allow levels of automation and can be quite efficient. This is also beneficial when considering the difficulty of sustaining consistent weeding action with a needs/judgment based model (time, variables, decisions, motivation), and with the belief that weeding should be sustainable, repeatable, consistent, and predictable.

Accordingly, qualitative methods should be primarily applied to acquisition and quantitative methods be primarily applied to weeding. Given this recommendation then, is there an opportunity for either methodology to find place in both? For the weeding algorithm proposed, the selector may need to pull a collection of titles before objective-based decisions are made. These are titles considered “untouchable” such as local authors, thesis/dissertations, local histories, classic texts, and so on. Although this type of resource is not too difficult to identify, for example, it either is or is not a local author, or thesis, etc., these decisions still require qualitative judgments on the part of the librarian. However, because qualitative judgments are applied in the preservation of resources, the professional is relieved of some of the conflicts inherent when this methodology is applied toward the removal of resources. Consequently, using this methodology is highly recommended for excluding important content from weeding prior to
applying an objective process. On the acquisition side, quantitative methods are beginning to find place in selection-on-demand schema. These provide powerful acquisition models that are based on patron wants and certainly provide predictors of future use. For the recommended weeding procedure, using a qualitative approach to exclude “untouchable” content prior to the algorithm’s use is the only instance of this methodology being employed in weeding. Its application is found in the sequence of algorithmic steps.

The Algorithm

Several variables can be used in a stepwise approach. One variable could be “frequency of use.” Use can be assessed in several ways. Search results, circulation statistics and in-library use should all be used when possible to create a usage frequency for the collection. Selth et al report from their study that 11% of books experience in-library use without any circulation. Including as many of these variables in the usage calculation is one of the ways the algorithm becomes more predictable. Slote’s shelf-time period can be used to determine percentage cutoffs and time periods. Otherwise, a simple percent of use over a predetermined time period (years) could be employed. For example, the librarian sets a time period of 15 years. A given title experiences 3 uses in that time, resulting in 20% use. Titles with less than 30% use would be eligible for weeding.

Another variable is “age.” Age, of itself, should not be a determinant in weeding titles. While it is true that older titles will likely experience less use, it is the frequency of use metric that governs these selections. There should be no discrimination based on publication age when calculating frequency of use. Studies have shown age to have little correlation here. If a title passes the frequency of use metric, it should remain regardless of its publication date. But another aspect of age to consider is the performance of newly acquired titles. Since usage is more reliably determined over a period of years, some titles may be ineligible for the frequency of use calculation. For example, if the frequency of use metric requires a time period calculation spanning 15 years, all titles that are 14 years or younger in age are effectively excluded. Should all titles within that range that have never been used remain for that long? Remembering Kent’s declining use percentages, newly acquired titles saw only 1-2% first time use by the 6th and 7th years. How many years will a newly acquired title remain without use before it becomes eligible for weeding? Trueswell recommends removing all books that have not circulated during the last 8 years. The librarian may find it useful to set an eligibility limit on newly acquired titles. To do this, a “new title age” variable can be employed.

Another variable is simply called “last use year,” meaning the period of time since the title was last circulated or used. Consider a title that fails the frequency of use metric. The use, from our example, falls below the 30% threshold. Now consider that the use, even though it was low, occurred within the last year. Is this a title that should be weeded? The “last use” metric prevents titles that have had recent use from being removed even though that use may be slight. It is up to the librarian to determine the cutoff period for this calculation. The question to answer is “How long can a title remain on the shelf without use before it becomes eligible for weeding?” The discipline considered will also demand differing cutoffs, as computer science will likely have a shorter cutoff limit than mechanical engineering.
In order to run the weeding algorithm, the following variables must be predetermined for each “total collection” (call number ranges/groups):

Time period = \( N \) = the number of years the frequency of use will be calculated.

Percent of use = \( p \) = the percent use cutoff under which titles are eligible for weeding.

New title age = \( A \) = the number of years before a new title is eligible for weeding.

Last use period = \( L \) = the number of years from the current year where any use preserves the title

Weeding Algorithm

The following terms are used:

\( Total\ collection = \) all titles in collection for a given call number range

\( Non\text{-}eligible\ collection = \) titles within the Total collection that will never be weeded (local authors, University thesis/dissertations, local histories, special collection titles, etc.)

\( Core\ collection = \) titles that will not be weeded (Non-eligible collection plus other titles as they are added by the algorithm)

\( Eligible\ collection = \) titles that are eligible for weeding (several modifications to this collection will be performed during the process)

\( Weeding\ list = \) the final eligible collection, titles that should be weeded

For each “total collection,” calculate the weeding list as follows:

1. Identify the Non-eligible collection and create the 1st iterations of the core and eligible collections.
   - This is the only step that requires a qualitative judgment. Titles selected here are those considered exempt such as local authors, University thesis/dissertations, local histories, classic texts, special collection titles, etc. This collection, once determined, can be used for future weeding evaluations. It requires updating only as non-eligible titles are added to the collection.
   - 1st core collection = non-eligible collection
   - 1st eligible collection = total collection – non-eligible collection

2. Apply “New title age” metric. Pull all titles that are \( \leq A \) years old since publication from the 1st eligible collection. Add this title list to the core collection.
   - 2nd core collection = 1st core collection + titles \( \leq A \) years old
   - 2nd eligible collection = 1st eligible collection – titles \( \leq A \) years old
3. Apply “Last use year” metric. Identify titles from the 2nd eligible collection that were last used during the most current range of years defined by $L$. For example, titles used/checked out within the last “$L$” years for Engineering are pulled and added to the core collection.

$$3^{rd} \text{ core collection} = 2^{nd} \text{ core collection} + \text{titles used within } \text{“}L\text{” years}$$

$$3^{rd} \text{ eligible collection} = 2^{nd} \text{ eligible collection} - \text{titles used within } \text{“}L\text{” years}$$

4. Apply “Frequency of use” metric. Run frequency of use on 3rd eligible collection. Compare percent of use over “$N$” years to “$p$” statistic. All titles with use $\geq “p”$ are pulled and added to the core collection.

$$4^{th} \text{ core collection} = 3^{rd} \text{ core collection} + \text{titles } \geq “p” \text{ use}$$

$$4^{th} \text{ eligible collection} = 3^{rd} \text{ eligible collection} - \text{titles } \geq “p” \text{ use}$$

5. The 4th eligible collection becomes the weeding list.

Conclusion

Due to conflicts, increasing variability, bargaining, political influences and other issues historically associated with applying qualitative methodologies during the weeding process, this paper proposes, with the noted exception of the non-eligible core collection, that qualitative methods be chiefly applied during the acquisition of titles. Qualitative methods are primarily assigned to the task of determining patrons’ needs since the methodology and variables match appropriately.

Quantitative methods are appropriately assigned to the task of determining patrons’ wants. Wants are calculated primarily by usage, so the variables employed are available only after titles have been made accessible to patrons. Consequently, patron wants are revealed when making weeding decisions. The algorithmic approach for weeding employs usage, last use, and new title holding periods in order to effectively cull collections of dead titles. Due to the volatility of disciplines and subject matter, collections should be built by combining call numbers with similar performance. These collections will each require adjustments to usage rates, last use, and new title holding periods.

References


20. Slote, 63.


22. McKee, 283-301.


29. Segal, 25.