

# UC Berkeley

## LAUC-B and Library Staff Research

### Title

On the Value of Shelf Reading: Exploring Methods Behind and Impact of Inventory Management

### Permalink

<https://escholarship.org/uc/item/9xw528fq>

### Journal

Technical Services Quarterly, 32(3)

### Authors

Mitchell, Erik T.  
Loo, Jeffery L.

### Publication Date

2015-06-15

Peer reviewed

# On the Value of Shelf Reading: Exploring Methods Behind and Impact of Inventory Management

Post-print of:

Mitchell, Erik T.; Loo, Jeffery L. On the Value of Shelf Reading: Exploring Methods Behind and Impact of Inventory Management. *Technical Services Quarterly*. 2015; 32(3):294–306.

doi:[10.1080/07317131.2015.1030270](https://doi.org/10.1080/07317131.2015.1030270).

Keywords: case scenarios, shared print, shelf reading, statistical techniques

## Abstract

In this article the authors explore questions of collection management and consider what methods libraries can use to evaluate the retrievability of an item from a stored collection. A case scenario method is employed to identify potential evaluation scenarios and specific quantitative techniques are considered that fill the evaluative need presented in each case scenario.

## Introduction

Libraries have long been concerned with ensuring accuracy and accountability for users who access collections, ensuring the long term durability of collections by mediating access through closed stacks or by imposing “no-go” sensors in open stacks. Although security is a concern, the broader ability of a patron to access an item by finding it on the shelf is the central focus of these efforts. To ensure this level of access libraries employ shelf reading programs in open stacks and closed stacks environments. Although libraries are trending more toward acquiring digital items, print procurement and management as well as the long-term durability of our print assets still requires considerable investment as libraries are not likely to dispose of these items anytime soon. At the same time, libraries are increasingly turning toward shared print programs and asking relevant questions around the number of copies required to support long term availability of items. This shift, toward seeking partnerships to ensure the long-term durability of physical items raises questions around how libraries ensure that the items are actually retrievable over time.

In this article the authors explore these questions in broad strokes by focusing on the notion of retrievability, what it means, and how libraries can put quality control measures in place to evaluate retrieval quality. In taking on this question is the authors are interested in exploring ways to ensure that the quality control (QC) measures are met. In doing so, they naturally focus on a specific type of QC—the process of shelf reading and the impact it has on measuring retrievability. This specific question is of interest given the new issues that are arising as print collections grow to very large scale and as they

become more geographically and organizationally distributed (e.g., regional and national shared print networks).

In order to explore this topic the authors consider approaches to evaluating retrievability broadly before considering selected case scenarios of retrieval QC and matching potential solutions. One underlying assumption of each of these case studies is that the QC measure is based on the match between metadata in information systems and the items on the shelf (e.g., shelving or circulation status). The authors explore one specific notion of retrievability—the question of whether or not an item that is expected to be on the shelf is actually retrievable. This question is a function of two related sub-questions: how accurate is the shelf-status metadata of a given item and how accurate is the order of items on shelves themselves.

## Literature Review

Retrievability is a broad issue in library materials access and the authors had difficulty finding a more suitable term to describe the act of retrieval based on the reviewed literature. Issues that impact retrievability include proper location on the shelf, inventory control via checkout and check-in procedures, unexpected inventory loss (e.g., theft), and unexpected inventory damage (e.g., fire or flood). Given each of these cases in which items can become unavailable, retrievability is the product of a complex and multi-faceted confluence of issues. As this article has already stated, retrievability is one of the often cited issues around preservation and shared print as programs grapple with the question of the number of copies required to ensure that an item persists for a given period of time (Yano, Shen, & Chan, 2013) and what the differences are in terms of preservation for shared print in place versus shared print in high-density storage approaches.

While libraries have automated inventory control mechanisms to deal with theft (e.g., “no-go” gates), loss (e.g., circulation records), and damage (e.g., item replacement policies and construction approaches) there are few “at-scale” methods to reliably answer the question “is the item retrievable from the shelf if the information systems register it as being there?” Radio Frequency Identification (RFID) systems employ metadata aware tags that can be read through radio sensors rather than via optical scanners, making it possible to scan books quickly without touching each one. RFID systems promised to be a step forward in simplifying inventory control through faster scanning of collections but the costs associated with converting from barcode to RFID inventory markers remain considerable, despite a trend in reduced costs for materials and infrastructure. For example, a 2006 (Singh, Brar, & Fong, 2006) review of RFID adoption referenced a cost of approximately \$1 per RFID tag to implement. Although RFID tag costs have declined considerably in recent years (e.g., a search for RFID labels shows prices from \$.20 to \$.75 per label), this represents a considerable investment for libraries with large collections (e.g., \$2.2 million for an 11 million item collection, not including labor and other infrastructure costs).

While there are a number of benefits from RFID technologies including easier inventory control, easier self-service around item circulation, automated sorting, and integrated security gate interaction, the cost of converting library systems and infrastructure to leverage these benefits adds to the overall cost of implementation. These features, taken together, represent a more complete approach to inventory

control and hopefully help increase levels of retrievability for items but the cost of realizing this improvement can be difficult to cover. In addition, simply detecting a missing item does not provide sufficient information for libraries to know whether or not the item is mis-shelved or missing entirely. This is such an important question for libraries we have a well-defined process and considerable investment in shelf reading practices.

### **The Cost of Comprehensive Shelf Reading**

The high cost of a comprehensive transition to RFID-based systems means that it is not likely to be the selected solution for many libraries. At the same time libraries invest considerable resources in traditional shelf reading, a process which remains little changed in the last 20 years. A 1979 article on shelf reading the two million volume collection of the Firestone Library cited the size of the collection and ineffective shelf reading methods as barriers to efficient shelf reading (Bennett, Buxton, & Capriotti, 1979). Bennett, Buxton, and Capriotti proposed a methodology for shelf reading that incentivized productivity (e.g., readers were paid by range read, not hours worked), reducing the time needed to complete an inventory from 4 years to approximately 771 hours of work. The shelf reading project in their study estimated that each section of shelving containing approximately 180–200 books took 14.4 minutes to read (24 employees \* 15 hours of work \* 60 minutes / 1500 ranges). Every 1,000,000 volumes (approximately 5555 ranges) would take around 1,333 labor hours of work. Since 1979, the collection of Princeton University, the collection on which Bennett, Buxton, and Capriotti worked, has grown to over 11 million volumes (<http://library.princeton.edu/libraries/firestone>). Even assuming that the process explored in this article was still the most efficient, Princeton library today would require 14,667 hours of work to inventory the collection.

This calculation of time assumes that there are no errors in shelf reading, does not take into account the amount of time spent organizing work outside of shelf reading, and does not accommodate actual work productivity schedules (e.g., number of staff to be hired to work and how many hours per day can be dedicated to this work). Using Bennett, Buxton, and Capriotti's (1979) measures (24 staff at 5 hours per day) an inventory of the 11 million volumes today would take 122 days of 24 employees working 5 hours per day but in all likelihood, libraries are not dedicating this level of staffing to shelf reading activities! Given this, such an approach would be difficult to sustain over time and would likely require continuous shelf reading.

Lowenberg suggests an alternative approach to shelf reading via representative sampling in her 1989 article (Lowenberg, 1989). Lowenberg's assertion—that shelving quality requires ongoing maintenance based on how frequently the collection is used—reduces the cost of shelf reading by requiring continuous reading of highly used areas only. Lowenberg suggests monthly reading of the top 20% of most used areas and in doing so seeks a way to mitigate the high cost of comprehensive shelf reading while delivering considerable value by keeping the most used collections organized. While this model may indeed have the benefit of improving retrievability it does not enable the library to make broad statements about the overall retrievability of the collection.

While many articles on shelf reading seek to design more efficient or impactful approaches to shelf reading it is less common to find research that evaluates the overall value of this work, particularly in

terms of the loss to collections or cost of alternative approaches. A 2009 article on shelf reading by Sung, Whisler, and Sung (Sung, Whisler, & Sung, 2009) evaluated the cost of shelf reading versus the theoretical cost of re-purchasing lost items that were “badly” mis-shelved (e.g., 25 items beyond the appropriate location). According to their research it took 707 hours to shelf-read 305,000 items, a rate of 25 minutes per “range” of 180 items. Sung, Whisler, and Sung estimated (based on a cost of \$30 per item in labor costs to replace an item) that the cost to replace items far exceeded the cost of shelf reading (\$159,000 to replace versus, \$11,000 to shelf-read for their collection) and made the assertion that regular shelf reading was an efficient use of resources. Posed with the question of scale (e.g., does the value proposition work as well at 11 million volumes), the cost of performing shelf reading for 11 million volumes using the Sung, Whisler, and Sung figures would be \$396,721. While it could also be claimed that the corresponding number of missing books would be found (e.g., 1.73% of items is equal to 191,000 items), such a program would require some data to confirm.

### **A Meta-Analysis on Shelf Order**

Bennett, Buxton, and Capriotti's (1979) work surfaced interesting numbers around the amount of mis-shelving in library stacks that contrasts with the overall finding of 1.73% of mis-shelved items in Sung, Whiler, and Sung (2009). Bennett, Buxton, and Capriotti found that books in some ranges were mis-shelved around 9% of the time (7) but did not provide an overall average because there were many sections without mis-shelving data. Overall, the range of mis-shelving was 0% to 9.7%. Their work also found that 96.6% of items that were mis-shelved were within two shelves of where they should have been, matching the findings of Sung, Whiler, and Sung. The authors recognize that shelf reading in a classification-based environment has the important outcome of improving retrievability. At the same time however, in high-density storage environments in which items are shelved by size, retrievability is a more pressing issue as mis-shelved items are not as easily found. While a meta-analysis of these shelving accuracy reports would be useful it is not possible to draw comparisons between the numbers as the selection methodology for each of the studies do not indicate that they used randomness or comprehensiveness to get a representative view of the collections being read. In order to get a real base-line idea of shelving order a more systematic approach to shelf reading via random sampling is warranted.

The authors recognize the actionable goal of many shelf reading projects (e.g., to ensure that each item is in order and retrievable). At the same time they approach the question of shelf location accuracy as a facet of collection risk and loss generally speaking. The goal of applying this lens to the issue of shelving accuracy is to ask questions around collection risk of loss broadly speaking (e.g., how confident can a library be that the items it thinks are available are actually retrievable?). These questions are applicable in shared print environments where libraries and high-density storage facilities seek to assess the risk to their collection without costly collection inventory practices.

In order to ask this question broadly, the authors make the assumption that shelving accuracy is a question of metadata accuracy (e.g., Does the item status marker in a metadata record match the actual status of the item on the shelf?) as well as shelving accuracy (e.g., Is the item where I expect it to be?). The implication of this approach is that trust around metadata accuracy can allow a library to make

commitments to shared print repositories without extensive inventory costs. Such an approach might even allow libraries and high-density repositories to use random-sampling inventory control techniques to make a collection retrievability assertion with a level of statistical validity (e.g., 99% available). Such an approach would be an innovative approach to allowing shared print programs to embed risk assessment into the question of “how many copies” need to be retained.

In regards to the open question of how reliable shelving status and order should be, a good starting place may be the work surveyed in this article. The lack of statistically valid approaches to selecting a subset of the population indicates that some baseline work is needed to discover what the overall expected reliability of shelving status and order should be. Granted, the question of metadata quality is but one facet of item retrievability and, on its own, is not sufficient for a library, high-density storage facility, or shared print cooperative to make an assertion about the number of needed copies to ensure long-term item preservation and availability. At the same time, the ability to make assertions about the reliability of shelving status metadata is a key enabler to the building of shared print collections as well as decision making around the risk posed by having items available in open versus closed stack collections.

## **Research Questions and Methodology**

Through this article the authors seek to answer two questions. First, what possible measures of retrievability are a good fit for overall collection maintenance questions? Second, how can we appropriately choose between these options? The first question is explored in detail in the literature review and the second question is explored via case scenarios. To review, measures of retrievability include metadata quality around shelving status, risk of loss due to poor inventory controls systems, risk of loss due to patron activity, and risk of loss due to large-scale facility issues (e.g., fire, flood). The authors recognize that a fully formed item preservation program needs mechanisms to track each of these issues (e.g., high quality inventory control systems, policies to require patron adherence to item use and return, and building management approaches to mitigate large-scale risk). At the same time, the authors make the assertion that a statistically valid sampling approach to shelving status questions allows libraries, high-density storage facilities, and shared print cooperatives to make valid assertions about the collections they manage at lower cost and higher confidence. In order to explore how an organization could approach sampling and surveying of collections the authors explore theoretical case scenarios that are typical to libraries, high-density storage facilities, and shared print cooperatives.

For each case scenario a brief “case” is introduced as well as potential methodologies. For the sake of simplicity a preferred solution is selected and explored in more detail including the benefits and issues with the approach as well as cost and benefit factors.

## Case Scenarios

### **Case scenario 1: A library has a small collection and needs to know which books to replace in its collection**

In cases where an organization has a relatively small collection or in which there is a desire to have data which is actionable (e.g., mis-shelved items are found and re-shelved) it is appropriate to employ a comprehensive shelf reading program. Techniques suggested by the reviewed literature include breaking up the process into easily managed components, making use of labor at appropriate levels (e.g., hiring students to do reading rather than having staff do the shelf reading) and making use of technology to improve process reliability and lower cost of information management.

The costs of such a process ranged from 750 books per hour (Bennett et al., 1979) to 431 books per hour (Sung, Whisler, & Sung, 2009). Both of these examples come from items shelved in classification order. A 2015 effort to complete shelf reading in a high density storage facility found that high-density storage shelf reading of selected items or of items outside of barcode order is considerably slower (e.g., 37 items per hour in a large storage facility).

The work to be done in complete shelf reading could follow a number of paths. It is most likely that shelf readers will be tasked with putting collections back in order as opposed to tracking items that are mis-shelved. In addition, unless readers are given a list to work from of items that should be on the shelf it is impossible for them to know which items should be present but which are missing. The advance of RFID and barcode scanning technology can facilitate this sort of data collection but can also pose issues for order-based shelf reading (e.g., a shelf-reader focused on scanning barcodes may not be able to focus on reading call numbers accurately). Likewise, good systems and real-time data are required to ensure that barcode collection or RFID reading systems are effective.

In cases where libraries are seeking to deliver a highly reliable shelving status it is possible to use a focused reading approach grounded in collection use or other measures as indicated by Lowenberg (1989). Such an approach can help libraries realize more impact from shelf reading work and potentially scale efforts in larger collections. At the same time however, the effort must be considered as a targeted effort that is very much a restricted implementation of the “full scan” approach. In other words, being intentionally selective about shelf reading specific parts of the collection does not result in data that can be generalized to make statements about the shelving status of the collection overall.

### **Case scenario 2: A library has a large collection and wants to be able to state a level of confidence around retrievability**

As the review of literature found, the costs of scanning large collections as well as the time required can have considerable impacts on the value of shelf reading in these cases. While libraries may be motivated to be able to make statements about the overall accuracy of shelving statuses in their collections they are unable to complete the shelf reading activity in sufficient time to make these statements valid. In these cases a library can use random sampling to select a representative selection of the collection for review.

By taking a random sample approach libraries can make assertions that the sample is representative of the population to a particular confidence level. In order to make such a statement a library must be able to take a random sample of their collection and check shelving status in a relatively short timeframe. In addition, while it is expected that workers should re-shelve items when they find them, they should also keep track of items that were mis-shelved so that they have a broad understanding of the shelving accuracy across the collection. In calculating sample size, a library must establish a confidence level (e.g., 99%) as well as a confidence interval (e.g., 1%) that they wish to adhere to. They can then use a sample size calculator (e.g., <http://www.calculator.net/sample-size-calculator.html>) to determine how many items need to be sampled to make a generalized assertion about shelving status.

### **Generating a random sample**

Once libraries have a sample size it is important to extract a randomly sorted list of items from the whole collection. This is most easily done by extracting the full collection in a list and by sorting in a random order before retrieval. This allows the library to easily expand the sample size if need be without starting from scratch with a new list of random records (e.g., if a library starts with a sample size of 100 items but later wishes to expand the sample size, they must work from the original randomly sorted list or they will have to discard the first 100 items).

There are many programmatic ways to approach random ordering of lists. Microsoft Access, for example, provides a simple programmatic means in queries to return a randomized list (e.g., the random function `rnd()`). When applied to a numbered field it returns the field in random order. By ensuring that the table of items in the collection includes an accession number (e.g., 1, 2, 3, 4, 5) and by applying a random function to it, the library can create a new table of the list returned in random order (<http://www.techrepublic.com/blog/how-do-i/how-do-i-retrieve-a-random-set-of-records-in-microsoft-access/>).

This approach can considerably reduce the number of items that must be checked but also complicates checking as items that may or may not be near each other. For example, being able to say that “our collection is 98% retrievable” for a collection of 11 million items requires a sample size of 16,616 items at a confidence level of 99% and a confidence interval of 1. Clearly the issue of item proximity for checking could be significant in an 11 million volume collection. For this reason it may seem preferable to break items into ranges following the model proposed by authors in articles reviewed in this article (e.g., randomly sample a certain number of ranges to check rather than checking individual volumes in that range). The problem with such a technique is that it greatly reduces the population (e.g., 11 million items could be grouped into 61,000 ranges).

In cases where libraries are randomly sampling collections or areas of the collection it is important to track the number of items that are mis-shelved or not findable. This percentage of items is the actual finding on the accuracy of shelving status. For example if 200 items out of the 16,616 items checked are mis-shelved or not findable then 1.2% of the collection is mis-shelved or not available (e.g., 200/16,616).

Being able to make such an assertion is valuable in a number of ways. First, it allows libraries, storage facilities, and shared print cooperatives to make assertions about the quality of shelving status in their



program or facility. This can enable decisions around how to allocate future resources (e.g., in more or less extensive shelf reading or in investment in replacement programs), how to evaluate the risk to a collection, or where further training or effort is needed in collection management. Making these assertions and selecting the right environment to apply such an approach should be approached carefully. For example, while this method is appropriate for statements around overall collection quality, it would not be a good surrogate approach for the collection replacement cost study of Sung, Whisler, and Sung (2009). Likewise, while it would have been an appropriate alternative approach for Johnson, Buxton, and Capriotti (1979), it would have required a completely different approach to staff allocation, training, and compensation.

At the same time however, shared print programs often struggle with the cost of holdings validation as part of the commitment and disclosure process. This is likely to become a more pressing issue as large scale shared print efforts become more common. Rather than accepting the notion that it is not possible to do holdings validation across the collection due to size, a random sampling technique could enable these groups to make assertions regarding the collection as a whole. This approach, when applied to the question of copies required for long-term preservation and access could be transformative for shared print programs. For example, if a shared print group were to simply aspire to have at least two copies of an item in a program it could effectively eliminate shelving status metadata as a risk to collection retrievability. Again, there are many other factors at play including inadvertent loss, theft, or damage, but in terms of being able to understand whether or not an item that we think is on the shelf is actually retrievable, a two-copy approach would provide extremely high reliability for retrieval.

### **Data gathering**

Because this approach is measuring the accuracy of metadata in regards to shelving status (e.g., marked as available and on the shelf, marked as available but not on the shelf, marked as unavailable but on the shelf, marked as unavailable and not on the shelf), it is important to capture each of these cases. In gathering the data it is important to recognize that there are two success cases (available and found, unavailable and not found) and two failure cases (available but not found, unavailable but found). If libraries are seeking to better understand the differences between types of materials, subject areas, types of stacks, or other differences among groups it is important to both generate a sufficiently large sample to include representatives from each group and also to gather any additional metadata required (e.g. the extent to which an item was out of order).

### **Descriptive and inferential statistics**

In this case scenario (e.g., statistical evaluation of a single group), the calculation of sample size is the most important statistical process. From this it is important to calculate descriptive statistics (e.g., percentage of inaccurately shelved items) at the group level only. Inferential statistical analyses could be conducted on this categorical variable on accuracy (i.e., accurate versus inaccurate).

The inferential test that is available for this dataset helps the library verify that the data is significantly different from a selected measure of metadata shelving accuracy (e.g., 98% accurately shelved). In this case, tests of significance (<http://www.stat.yale.edu/Courses/1997-98/101/sigttest.htm>) and confidence

intervals (<http://www.stat.yale.edu/Courses/1997-98/101/confint.htm>) for a single proportion would use a z statistic.

### **Case scenario 3: A library has collections in multiple locations and wants to compare the shelving status reliability of these locations**

While the previous methodology approached shelving quality assuming a single location, this case scenario examines shelving quality issues from multiple locations or for multiple groups. It is important when engaging in statistical sampling techniques to not conflate populations when you actually need to answer questions about different facets of the population. This may occur for libraries seeking to identify the collections that are at most risk of being out of order, for high-density storage facilities seeking to evaluate the relative retrievability of collections from individual storage facilities, or from shared print groups trying to evaluate the retrievability of open versus closed stacks.

In order to complete this comparison, it is important to track each sub population as well as treating the group as a whole. For example, if you have two storage facilities and are seeking to evaluate the retrievability of each facility you must make sure that you evaluate a statistically valid sample from each facility in proportion. For example, facility A has 1 million volumes, facility B has 2 million volumes. Overall, the two facilities must check 16,550 items to make a statistically valid assertion about collection retrievability but they must do so in proportion (e.g. facility A should check 5,516 items and facility B should check 11,033 items). The reader should consult the second case scenario for a method of generating randomized lists.

#### **Descriptive and inferential statistics**

In the case where libraries are seeking to compare two groups (e.g., comparisons of shelving metadata accuracy across two locations or open versus closed stacks) a contingency table should be created to understand the distribution of accurate versus inaccurate records across the types/locations of storage facilities. Such a contingency table is represented in Table 1.

Table 1. Contingency Table for a Study of Shelving Accuracy in Multiple Locations

	Accurate	Inaccurate
Facility A	999	1
Facility B	998	2

A chi-square test (<http://www.stat.yale.edu/Courses/1997-98/101/chisq.htm>) could then test the associations between the accuracy and location/type variables. Stated another way, chi-square tests allow the researcher to check whether or not the difference in shelving metadata is significantly different between facility A and facility B. This could be helpful for benchmarking performance and for identifying associations between storage procedures and accuracy rates. A suitable online chi-square test for this type of test is at socialstatistics.com

(<http://www.socscistatistics.com/tests/chisquare/default2.aspx>).

In the example provided in Table 1, the calculated chi-square statistic is .3338,  $p = 0.56341$ . Therefore, the difference in metadata shelving accuracy between facility A and facility B is not statistically significant. Finding these measures of significance are important in understanding what the sampled data is indicating about collections.

## Discussion and Conclusion

The three case scenarios examined in this article suggest appropriate times to employ full-population versus random sampling techniques to evaluate the reliability of shelving status metadata. One of the key issues explored in the literature review as well as the case scenario is that randomly sampled approaches are best suited for quality of service or collection reliability measures and are not particularly well suited for “actionable” results. While this is an important distinction to make it is also true that libraries with extremely large collections may want to use random sampling as a way to check some subset of a collection (e.g., items most circulated, areas most browsed).

One of the key issues with moving to a random sample approach is that the time required to complete a shelf-scan of items that have been randomly selected is much greater than comprehensive scans. This can be mitigated by the ordering of selected items by classification or location ordering but it is not possible to achieve the same efficiency as full scan approaches. In fact it is worth considering the cost of selective versus full scans for smaller collections given the number of items that must be scanned for small and medium sized collections. Another issue the literature review has raised is that currently, there is no established baseline of accuracy at storage facilities. Should we say retrievability is 100% or is that unreasonable? What is the established baseline for accuracy? Should we do this study nationally to find the average accuracy rate for a baseline? The shelving accuracy data surfaced in the studies consulted varies considerably and, given a lack of statistically valid sampling techniques, it is difficult to know if these shelving accuracy rates are representative.

In this article, the authors have proceeded with the assumption that metadata accuracy around shelving status is a valuable measure of collection presence and availability. The example pursued finds the most value for shared print programs seeking to make assertions about collections as they are committed to a program in the absence of a full validation of holdings. This may be an approach with acceptable risk (e.g., particularly in cases where multiple copies exist or in which print and digital copies exist) but may not be appropriate in all areas. In order to understand the implications of random versus comprehensive collection analyses more work is needed to actually apply these techniques in real world settings. In addition, the authors focus on relatively simple statistical analyses of collections. It is entirely possible that other statistical analysis techniques could yield greater efficiency.

## References

1. Bennett, M. J., Buxton, D. T., & Capriotti, E. (1979). Shelf reading in a large, open-stack library. *The Journal of Academic Librarianship*, 5, 4–8.

2. Lowenberg, S. (1989). A comprehensive shelf reading program. *The Journal of Academic Librarianship*, 15, 24–27.
3. Singh, J., Brar, N., & Fong, C. (2006). The state of RFID applications in libraries. *Young Adult Library Services*, 4, 24–32.
4. Sung, J. S., Whisler, J. A., & Sung, N. (2009). A cost-benefit analysis of a collections inventory project: A statistical analysis of inventory data from a medium-sized academic library. *Journal of Academic Librarianship*, 35, 314–323.
5. Yano, C. A., Shen, Z. M., & Chan, S. (2013). Optimising the number of copies and storage protocols for print preservation of research journals. *International Journal of Production Research*, 51, 7456–7469.