Charleston Library Conference

Accidental Collection Assessment: The NCSU Libraries Collection Move

Annette Day
North Carolina State University, annette_day@ncsu.edu

John Vickery
North Carolina State University, john_vickery@ncsu.edu

Hilary Davis
North Carolina State University, hilary_davis@ncsu.edu

Annette Day, John Vickery, and Hilary Davis, "Accidental Collection Assessment: The NCSU Libraries Collection Move" (2012). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284315092

Accidental Collection Assessment: The NCSU Libraries Collection Move

Annette Day, Head, Collection Management, North Carolina State University Libraries
John Vickery, Collection Manager, Management and Social Sciences, North Carolina State University Libraries
Hilary Davis, Associate Head, Collection Management and Director of Research Data Services, North Carolina State University Libraries

Abstract

In January 2013, the NCSU Libraries will open a brand new library, the James B. Hunt Jr. Library. Approximately 1.4 million volumes will be transferred to the new library from existing campus locations for opening day, with the majority of those collections being housed in an automated retrieval system (bookBot) and not in open stacks. A collections move of this size requires strategic planning reinforced and guided by a lot of data, assessment, and clean-up of records, items, and processes. This system-wide effort has given us unique and unprecedented opportunities to assess our collections, their scope, access, composition, and trends in growth and use.

Introduction

This paper will describe how this move provided us with unexpected opportunities to assess our collections, their scope, access, composition, and trends in growth and use. We'll describe the move process and what data we needed to gather to plan for the move. We'll also provide some case studies to illustrate the types of analysis undertaken.

Hunt Library Background

Hunt will be a second main library focusing on engineering and computer science disciplines. NCSU's other main library, the D. H. Hill Library, will focus on science, humanities, and social science disciplines. The library system also has three branch libraries supporting veterinary medicine, design, and natural resources. These libraries support a student body of nearly 35,000 members and a faculty community of close to 8,000 members.

The Hunt Library project was initiated in 2007, and a little under 6 years later the Hunt Library will open its doors. Hunt Library will be located on our Centennial campus. Centennial campus is a unique community of NCSU colleges, research units, and corporate and government partners. Until now, the campus has been lacking a signature focal point, a place for these different constituents to interact, collaborate, and innovate. Hunt will provide that focal point.

From the start, plans for the Hunt Library were ambitious, and our goal has been to create the "the best learning and collaborative space in the country." Hunt will support the disciplines that are located on Centennial Campus: engineering, computer science, and textiles. The Hunt Library is integral to the NCSU Libraries' vision to be NCSU's competitive advantage. We want to be a library that is a major factor in attracting and retaining the best faculty and students.

The library will be a technology incubator providing students and faculty with access to the latest technology in its spaces. The ambition is to prepare NCSU students to lead and support cutting-edge research in a technology-driven economy. In addition to the basics that students expect in today's libraries such as consistent wireless connectivity and flexible, easy, copious access to electrical power for charging their devices, the environment in Hunt Library will be filled with virtual browsing, video-conferencing, online room scheduling, collaborative projection, whiteboards, and video walls.

Creating flexible and usable learning spaces that meet the needs of diverse user groups is obviously

http://dx.doi.org/10.5703/12882843150792

a central and key goal for Hunt. The building will recognize the various needs of graduate and undergraduate students, faculty members, and researchers and will offer spaces tailored to those unique communities, including a Research Commons for faculty and a Graduate Commons for graduate students. Both of these spaces will provide restricted access solely for those constituencies so that we can ensure dedicated space in the building for them. There will also be a Skyline Reading Room, a Creativity Studio, and a Teaching and Visualization Lab, as well as many group study rooms and individual study seats. There are over 100 bookable group study rooms in the building. One of the most consistent pieces of feedback we've received from students over the years is that they need more collaborative spaces, and Hunt will provide that.

Currently the most heavily used building on campus is our other main library facility, the D. H. Hill Library. This frequently has more than 10,000 visits in a day and is struggling to accommodate the demand for study and collaborative space from campus. Students are often unable to find seats. Ranked lowest in the University of North Carolina (UNC) system in terms of square feet per student, NCSU Libraries are able to provide seating for less than 5% of the student body, while the UNC standard is 20%. The Hunt Library will double NCSU Libraries' available study seats to help narrow this gap.

The Hunt Library will house print collections in an automated book delivery system which NCSU has named the "bookBot." The bookBot can hold up to 2 million volumes allowing us to free large amounts of space in the building which we've dedicated to technology-rich study and collaboration spaces.

Hunt Library Collections—bookBot

There are two collection types that we needed to consider in the move: those going into the bookBot and those going into the open stacks. We will describe the bookBot collection first.

The lion's share of materials transferred to Hunt Library has been or will be going into the bookBot. It will primarily hold the working collections of engineering, computer science, and textiles, but it will also serve as general storage for materials across the library system. Older and low-use materials such as journal backfiles, low circulating monographs, and other low-circulating collection types such as government documents and microfiche will all be housed in the bookBot. The majority of these low-circulating collections were transferred from off-site storage locations, and the bookBot is now our primary storage facility. Other materials were identified from the stacks collections of branch libraries with the aim of freeing up space. Journal backfiles gave everyone the biggest "bang for the buck" in terms of freeing up space. We undertook a large project to identify titles that had stable electronic counterparts and moved the volumes covered via online platforms to the bookBot. It was important to try to co-locate journals and not split them across different locations, for example, some volumes in a branch, some at Hill, and some in the bookBot. In addition to general collections materials, some special collections materials will also be stored in the bookBot because it provides controlled environmental conditions with no leaks or humidity.

From the perspective of collection management librarians, there was a strong desire to move as much to the bookBot as possible. One main benefit to adding any materials to this location is that it will provide us with much more clear usage pictures. For anything to be retrieved from the bookBot, it needs to be requested, so we'll be able to monitor use and perhaps move items to open shelf locations or weed them for good. Although the bookBot can hold 2 million volumes, we plan to move 1.4 million volumes by opening day and use the remaining capacity for growth.

The technology underlying high-density automated shelving has been used in large-scale industries for many years (e.g., UPS and Nike use automated retrieval systems). In recent years the technology has been adapted for use in research libraries. In a nutshell, the bookBot is a system where barcoded items are stored in barcoded bins. The bins are of differing size to maximize capacity and to provide the best environment for the materials. Every item that is to be transferred to the bookBot has to be measured so we can assess what size bin it needs to go in and plan the move accordingly. To ingest an item into the bookBot, the item's barcode has to be scanned into the bookBot's inventory system as does the barcode of the bin it will live in. That way the system knows where it is. When we want to retrieve an item the bin is recalled and staff will manually pull the item from the bin and scan the item so we know it's no longer in the bookBot. On return to the bookBot the item can be added to any bin; it just needs its barcode and the bin barcode scanned into the system again. It's a simple system, but to enable it to work for us every single unit going into the bookBot needs to be barcoded. One of the very first steps involved in this move was a mass barcoding project.

From the patron perspective, requesting a book should be straightforward. They request items through our catalog using familiar screens they already see for recalling checked out materials or placing interlibrary loan requests. We estimate about a 10 minute turnaround time from point of request to getting it to the patron. This will depend on the time of day and staffing levels, but that is currently our service goal.

The NCSU Libraries are concerned that patrons may miss the serendipitous opportunities that physically browsing a collection brings, so we have created a virtual browse interface that includes tables of contents for the entire collection, not just items in the bookBot. The added benefit of the virtual browse environment is that it pulls together both physical and online materials in the same interface so the patron gets the full picture of the collection, something they can't do while physically browsing the shelves. In addition we are focusing on services to deliver collections to users to alleviate their concerns about reducing browse-able shelves. We plan to provide office delivery for faculty for items being requested out of the bookBot and will provide scanning services for journal articles and book chapters.

Hunt Library Collections—Open Shelves

In addition to the bookBot there will be some open shelving at the Hunt Library which will hold approximately 28,000 volumes. The collections in these spaces will primarily be current monographs and journals with some "niche" browse-able collections including faculty publications, classics in engineering, and fiction. The open shelving spaces at Hunt are dispersed and small, so these niche type collections fit the spaces well.

Goals of Move

Because every item going into the bookBot requires a barcode (and not everything has a barcode), the bookBot adds an additional layer of complexity into the move process. To add to the complexity of the move we are moving a significant portion of our collection to the bookBot—1.4 million volumes—in a truncated timeline of just 5 months. This means that we needed to plan for enough time to physically prepare the collection for moving, including adding barcodes to items that didn't already have barcodes, and we needed to define all issues and workflows before the move commenced.

We had three main goals for the collection move: minimize "downtime" of the collections, minimize patron disruption, and to use a transparent communication plan.

To minimize the downtime of the collections, this meant that we needed to make sure that the library catalog was as current as possible. During the course of the move, item records had their locations switched the night before the items moved next day. We also made sure that patrons could always use familiar processes to request any item (even those on the move) and that we would pull any requests from the move process and get them to patrons as quickly as possible.

To minimize patron disruption, we needed to make sure that movers and staff working to prepare the materials used plastic sheeting to limit dust and kept noise to a minimum. Movers were equipped with backpack vacuum cleaners to collect dust, and they worked within small spaces so that shelves were accessible much of the time. This also meant that we worked to ensure that no spaces were off limits to patrons. All study spaces and computer labs were available at all times during the move. In terms of timing, the move started in summer 2012 and will continue up until mid-November. The move will temporarily halt during exams and will be completed in mid-December after exams.

To achieve a transparent communication plan, we targeted our messages to focus on the practical issues and benefits of the move to our patrons. A "microsite," a website with targeted messages for staff and patrons, was created that included a daily calendar for when, why, and how the collections were being relocated as well as contact information and an FAQ. Physical signs were placed at the entrance of the library and on all floors so that patrons would know what was happening on a daily basis. We also leveraged our social media outlets (Twitter, Blogs, Facebook status updates) to convey the plan for the move and how it progressed day to day.

Staggered Move Strategy

For a move of this complexity, we used a staggered move strategy. The collections were moving from multiple origination locations and were going to multiple destinations (not just the bookBot, but also to other existing library locations). Because much of our collection needed to be barcoded, the staggered move strategy included building in time necessary to add barcodes, fix records, and repair physical conditions of items. The staggered move strategy also meant that the last collections to move would be those that will be going on open shelving: books published from 2007 to present, reference, and showcase collections.

Scope of Collections Move

All along the way, we needed to paint a picture of everything, where it was and what it was, getting more and more granular along the way. At the beginning of the planning process for the collection move, we were focused on the bird's eye view. To first determine how to redistribute the collections, we mapped call numbers mapped to those academic departments that we knew would have a home near Hunt Library. Early on, we really only focused on the broad strokes so that we could estimate the scale of the move based on the kinds of collections we were dealing with (including government documents, books, serials, maps, special collections, theses/dissertations, and various media) and the preparation needed in terms of staffing and time to prepare the collections to be moved.

As we got closer to the start of the collection move, we had to generate more granular data about the collection. We needed to get actual counts of items in the various locations across the library system for requesting move-bid estimates and determining fit on shelving (for each collection). We needed to get a handle on the physical parameters of items so that we could estimate the number of different sized bins to order for the bookBot. We needed to track the status of items so that we could make sure that checked-out items would go into the preparation workflow upon discharge, and so that we could get a realistic count of missing items to fine tune space allocation. We also needed to get a full inventory of items and their barcodes to coordinate the work of changing the locations of items and ingesting the physical items into the bookBot the next day.

Bumps in the Road

There were some problems that we did anticipate such as large chunks of the collection that needed to be barcoded. But there were some unforeseen issues that we encountered when the collection move got underway. There were items that we thought were missing but weren't (and vice versa); there were items that were in the wrong locations which meant that we had to hunt many items down (time consuming and labor intensive); and there were a mix of old and new practices that resulted in a lot of extra work to homogenize large chunks of the collection, such as the way accompanying materials were handled so that they could be requested out of the bookBot and across all library locations.

Data Strategies

Early on, we were more concerned with snapshots and bird's eye views, so we ran reports in our ILS (Sirsi Workflows) and imported them into spreadsheets to facilitate discussions and decision making. But as we got closer to the move, we needed to switch gears to get more detail at the item or barcode level. We needed to analyze the myriad ways of slicing and dicing the collection so that we could experiment with the distribution of the collection within the open shelving possibilities and within the bookBot across the different kinds of storage bins. We also needed to be able to answer questions like: Where do the collections really need to be located based on use patterns? Do we need to provide better access to these distributed collections by adding duplicates or buying e-books? How fast is the collection growing in print (considering trends in e-content growth, budget fluctuations, publication trends, etc.)?

A little under two years ago, we switched to a more sophisticated analysis mode using SAS software which made it much more manageable for us to do the kinds of analysis we needed to undertake. The following sections provide an overview of how SAS worked for us and include specific-use cases that were particularly well-suited for using SAS.

Scaling Up Our Analytics

In the following section, we discuss how we scaled up our analytics by using SAS programming. We also show examples of the type of analysis we've been able to do.

At the NCSU Libraries, the move planning and analysis became too complex to rely on the canned reports from our ILS and Excel for data analysis. While we do still use our ILS reports and Excel, we weren't able to answer questions as effectively as we wanted to prior to using SAS® software as our analytics tool.

What Is SAS® Software

We often think of SAS® software as a statistics package, but it is much more flexible and robust. SAS® software is an integrated system of software for business intelligence, forecasting, data warehousing, statistical analysis, and many more specialized functions. SAS® software is typically a programming based system. That is, code is written and programs are submitted to process and analyze datasets. While writing code is typically necessary, we believe the learning curve is worth the time investment to learn. SAS also provides a graphical user interface which allows a point-and-click environment to the programs running in the background. New users can even access the generated code to begin learning SAS programming.

SAS® software is also a licensed (not free) product. However, in the academic environment, it is often licensed at the university level similar to other software products and may not even involve an additional cost to the library. We encourage libraries to investigate available licenses that may permit use of the software.

Why Choose SAS® Software

There are many other tools available you could use for the analysis that we have done, but we believe that SAS® software is one of the best. Perhaps it is even the best. It is a tool and programming language specifically designed for data analysis and manipulation. Here, we look at five reasons to choose SAS® software.

1. You are in control of how you work with your data.

The programming language is flexible with concise procedures that encapsulate analytical functionality. The data manipulation functionality allows you to transform, recode and prepare your data for analysis.

Any data analysis project often generates additional "what-if" questions. Because SAS® software is programming based (or stored processes based if you use Enterprise Guide), you can easily adjust and repurpose your programs to re-analyze your data in new ways.

While there is a steeper learning curve to SAS than with tools such as Excel and Access, the payoff is worth the investment. Learning just a few data manipulation (DATA Step) techniques and procedures (PROC Step) can greatly extend analytical capabilities.

2. You can read data from any source. The programming language allows you to import data regardless of format. This includes the raw output from your ILS ranging from the

bibliographic tables to the transaction logs. SAS can also easily read data from Excel.

3. Big data is no problem. We're able to easily work with millions of records, and we could scale that up even more if needed. Our typical tools, Excel and Access, begin to show serious strain in terms of performance and capability well below this threshold.

Keeping the Move in Sync

In order to minimize collection downtime and the amount of time collections were listed as "in transit", the move of physical items had to be kept in sync with nightly data loads into the bookBot inventory system. This process required several steps as we cycled data from our ILS system through analysis using SAS® software and then

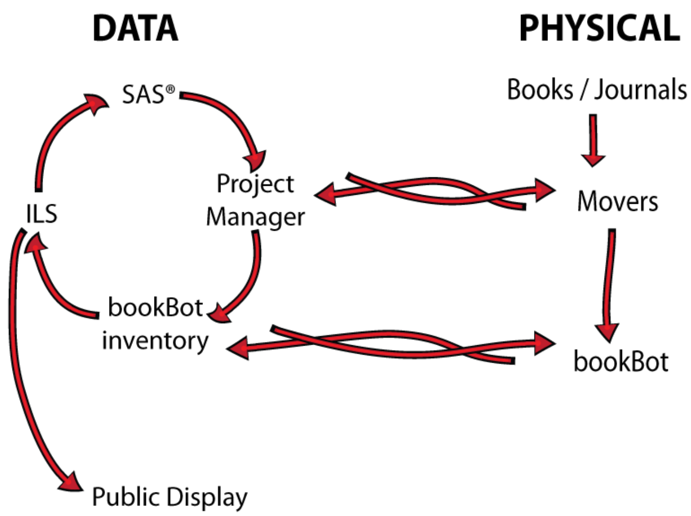

Figure 1. Syncing the Move of Data and Physical Items

4. Robust and customizable reporting functionality. Much of what we do is take millions of data records and distill them down to a report or picture that an administrator can use to make decisions. SAS® software can generate reports in many different formats and styles such as HTML, PDF, XML for Excel and RTF for Word. Multiple types of output can also be generated simultaneously from the same analysis meaning you can target more than one audience at a time.

5. Visualization functionality. This goes hand in hand with the reporting functionality. SAS® software procedures can turn your data into high quality, easily interpretable charts and graphs.

into human-readable form where our project manager could decide on a daily basis which records from our catalog needed to be loaded into the bookBot inventory system. Figure 1 illustrates this process.

Planning the Move

Analyzing the data using SAS® software allows us to accurately pinpoint the number, location and move criteria for each item in our collection. We are able to start with a dataset of more than 3.5 million records and pinpoint a single item.

Figure 2 shows a portion of an Excel file generated from a reporting procedure to identify and count each item's move status. Circled in red is an example of how, by using SAS® software, we can identify even singular items.

A	B	C	D
Number of items to be moved to HUNT Library			
Items from SATELLITE			
		Hunt Status	
Criteria for Hunt move	**Item Type**	**No**	**Yes**
		38,127	*735,798*
No criteria for move	MANUSCRIPT	10	.
	MAP	24,118	.
No criteria for move		*24,128*	.
Withdrawn Design items	AUD-CD	3	.
	BOOK	8,484	.
	CD-ROM	128	.
	SERIAL	4,716	.
	SOFTWARE	1	.
	SRLCIRC	607	.
	VID-CASS	53	.
	VID-DVD	7	.
Withdrawn Design items		*13,999*	.

Figure 2. Pinpointing Each Item in the Collection

Estimating Use and Impact

The following examples were in response to questions we needed to answer for our administration. This particular example demonstrates how we used SAS® software to analyze Sirsi transaction log data. We needed to address faculty concerns about the availability of items during the move and we also needed to think about how many people we would need to

This example also illustrates how SAS® software is particularly good at reading in “ugly”, non-standard raw data. Figure 3 shows two example records of the data. Circled in red, we have multiple types of delimiters. Highlighted in red, we can see that data values are tagged with data codes instead of being in regularly arranged fields. The records are also of varying length. This data cannot be simply dropped into Excel.

```
E20120711105816O073R
^S81JZFFCIRC^FESATELLITE^FcNONE^FWMDWALL
AC^UkZLWHITE^NQS01285303M^IQM1500 .W13
W35 1991
v.2^IS1^DHNO^HB1/7/2013^HELIBRARY^HFN^HI
N^HKCOPY^HODHHILL^dC3^Fv3000000^^O|UND
ERGRAD|NDS|NCSUSTAFF||||

1207021249290010R
^S82CVFFSTANDALONE^FEDESIGN^NQS02421189
R^OM^Fv200000^DB20120702124929^^O00084|
GRAD|ID|||||
```

Figure 3. Ugly Data

staff the bookBot stations where items are actually taken out of the big bins. We used the Sirsi transaction logs to produce analytic output that could answer the above questions and inform decisions.

After processing and analyzing the data using SAS® software, we were able to estimate use and impact of the bookBot. With output in such as Figure 4, we’re able to show both peak times and what class of users will be most affected. We can also look at that down to the hour of day.

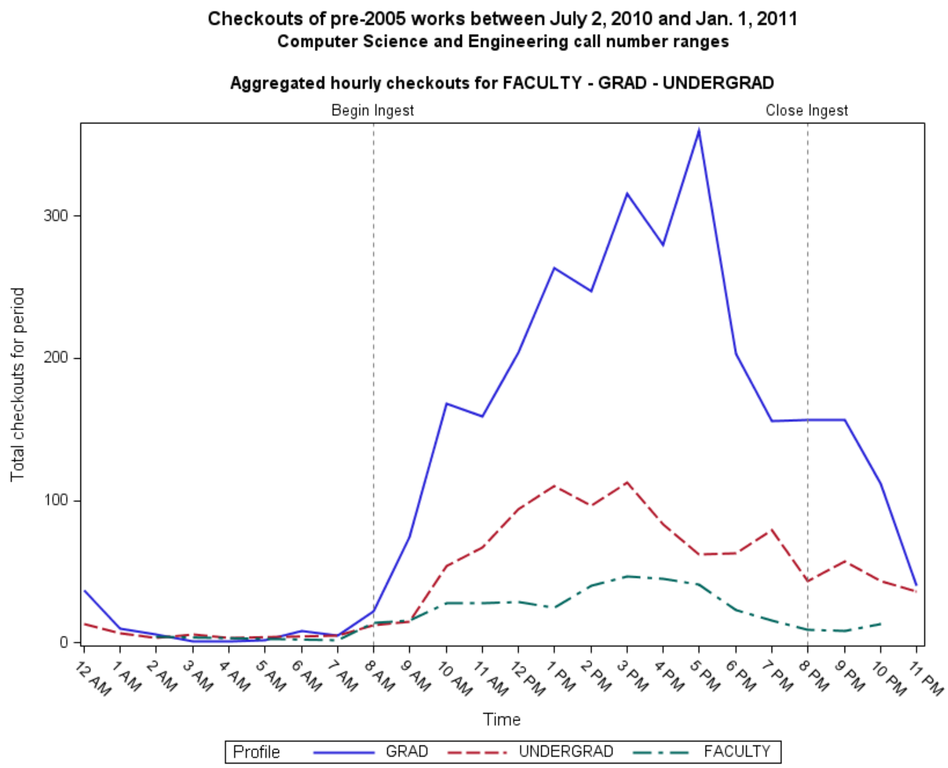

Figure 4. Estimating Use and Impact

We were also able to get very specific as we planned to staff the bookBot workstations. Figure 5, shows an example of box plot output that we used to show demand by hour for any particular day of the week.

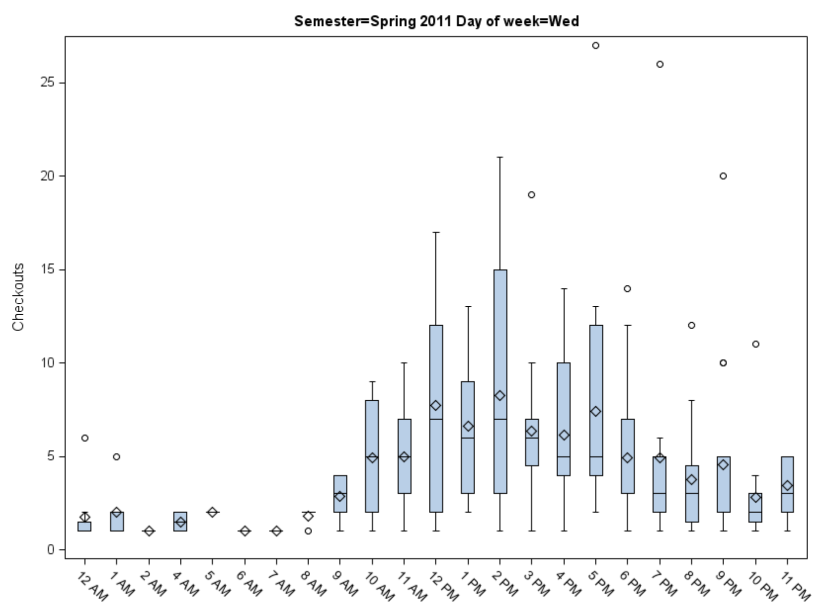

Figure 5. Box Plot Example

Forecasting Future Growth

Following is an example of how we applied modeling techniques to forecast future growth of our print collections that would not be moving to the new Hunt Library and bookBot.

The move of approximately one third of the D. H. Hill collection to the Hunt Library will free up a space in the D. H. Hill library, and we wanted to answer questions such as: How fast are we going to fill it up? How should we arrange the stacks now to minimize the chance that we'll need to do a major stacks shift in the near future?

In order to answer these questions, we used SAS® software to simulate our historical additions excluding those which would have been added to the bookBot based on their subject classifications. We then translated the number of items added in any year to the amount of physical shelves they take up on a cumulative basis. Finally, we used a modeling technique to forecast historic additions to the collections 5 years into the future.

Figure 6 shows how we can visualize the extent to which the D. H. Hill library will be at capacity. The blue line plots the actual values and the red are the modeled forecast values. L95 and U95, the green and brown lines respectively, plot the lower and upper 95% confidence limits. While not shown here, we can also drill down to any particular subject classification and floor combination to allow for stacks managers to better plan how to distribute the collections across library floors.

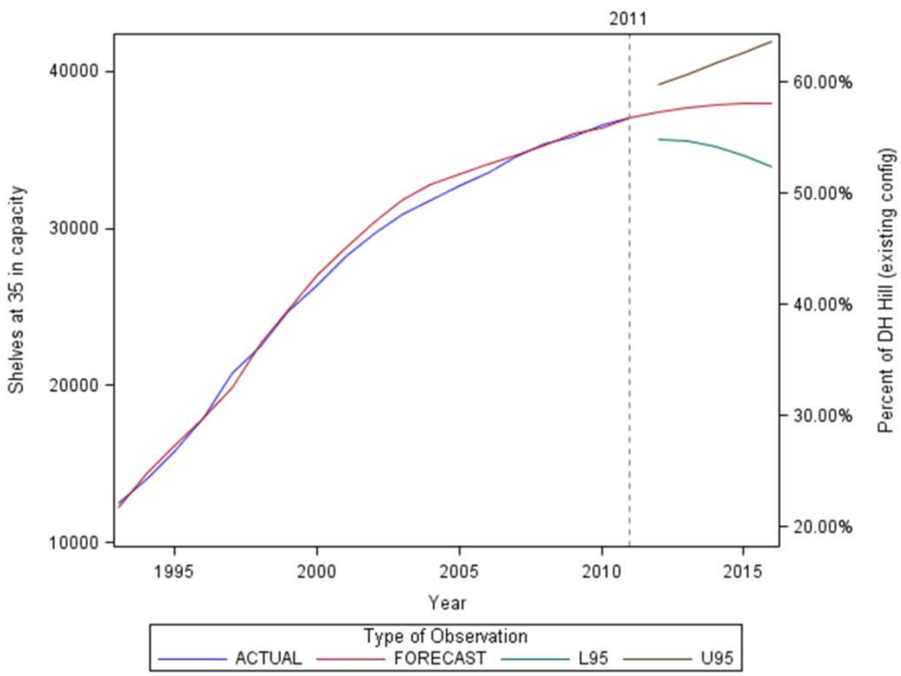

Figure 6. Forecast Example

Conclusions

Obviously a move of this scale is challenging, as there are many moving pieces involved. Planning is essential. Understanding what questions you need answered and leveraging the data you have available to answer those questions is fundamental to moving forward successfully.

The move planning and data analysis described in this paper was not something we did in isolation in the Collection Management department. It was vital to collaborate across many departments and groups including the main Project Manager for the move, Library Information Technology, Metadata & Cataloging, and Access & Delivery services to articulate our questions, assess our data needs, review current practices, and actually get the items moved.

Gaining the appropriate skillset to process the data was vital to us being able to answer the questions on the table. Our ability to work with SAS has been a somewhat serendipitous confluence of events, but it has opened our eyes to the value of having such a skillset in the Collection Management department. It has become an integral goal of the department to improve the data skills of our staff, as we believe it is a valuable investment of time that will pay dividend in the future. Collections librarians need to play a key role in analysis efforts as they understand the collections, their composition, how they have been built over the years, and how that is changing. We are in the best position to answer questions about our collections, and we need to make sure that our staff is in possession of the skills that allow them to do that.

Slides for this presentation can be found at http://www.slideshare.net/hilarymdavis/accidental-collection-assessment.