Collection Intelligence: Using Data Driven Decision Making in Collection Management

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

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

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Annette Day and Hilary Davis, "Collection Intelligence: Using Data Driven Decision Making in Collection Management" (2010). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284314822>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

COLLECTION INTELLIGENCE: USING DATA DRIVEN DECISION MAKING IN COLLECTION MANAGEMENT

Annette Day (annette_day@ncsu.edu) *Head of Collection Management, North Carolina State University Libraries*

Hilary Davis, (hilary_davis@ncsu.edu) *Associate Head of Collection Management, North Carolina State University Libraries*

Today's collections librarians are all too familiar with the harsh decisions involved in building relevant and quality collections. We see our buying power diminished by the woes of shrinking, or if we're lucky static budgets, and inflation increases. The recent economic climate has exacerbated this situation so that libraries are now facing unprecedented budget cuts. Yet these deep budget cuts do not align with the demand we are seeing for our library collections and services. Gate counts are rising, patrons expect the library to be open 24 hours and researchers are demanding access to the new journals and digital products that are launched each year.

With increased demand and reduced budgets we are left with the very difficult balancing act of deciding what we can afford and what we can live without, while still striving to provide quality support for learning, research, and teaching. As we attempt to stretch our diminishing budgets as far as we can and get as much value for our dollar, we need to be able to justify our decisions both internally to library administration, and externally to campus. For collections librarians we need to show that we are building relevant and timely collections that support the research and instruction needs of campus, whilst leveraging efficiencies that resource sharing and technology may offer. We need to show our collection intelligence, because if we cannot substantiate our decisions, the library may lose credibility as an effective partner and responsible steward of the collections budget.

So how does a library demonstrate its collection intelligence? At the NCSU Libraries we practice a data-informed approach to collections decisions, making use of data about our users and about our collections to inform our actions. We are not unique in this practice. Many libraries have placed heavy reliance on use metrics such as journal full-text usage statistics, circulations, gate counts, and interlibrary loan (ILL) transactions to help guide tough decisions. Specifically for collection management decision-making, there is an array of data that can be used to support your decisions and advocate for your collection. These include: cost of books/journals/databases/videos, etc., usage (full-text downloads, circulations, ILL transactions), rights (owned/leased), citation and publication patterns by your campus researchers (i.e., which journals are they citing and which journals are they publishing in), journal impact factors (as flawed as they are, they are still seen as a benchmark among many researchers and libraries), regional holdings, and journal editorial activity of your campus researchers.

Once you have this data there are many ways to work with it to support and advocate for your collections. This paper will describe three projects undertaken by collection managers at the NCSU Libraries to show how we successfully used a portfolio of data, metrics and tools to demonstrate our collection intelligence.

JOURNAL CANCELLATION PROJECT

Project Overview

At the start of the fiscal year 2009/2010 the NCSU Libraries was facing a cut to the collections budget of 15% or \$1.5 million dollars. To tackle this deficit we had to take several measures, the most substantial being a serials review and cut. The review was planned in consultation with the University Library Committee, a standing university committee that advises the library on policy and programming, and our faculty and graduate library representatives, appointed representatives from each academic department that act as formal liaisons between the library and the campus. By involving these groups in the initial planning stages of the review, we collected important feedback about the best way to design and communicate the review process to maximize input.

One key piece of advice we received from our advisory groups was that we should not present every single journal subscription to campus for feedback, but instead provide them with a shorter list of potential cancellations. So the review process began with collection management librarians reviewing all of our journal subscriptions and putting together a list of lower use (in terms of full-text downloads), less relevant, titles that we would present to campus requesting their feedback. This list consisted of 1,112 titles. It should be noted that the qualifications of “lower use and less relevant” were somewhat relative at this stage as these cuts were making a significant impact, but we wanted to avoid asking campus to review titles that we knew were high use and highly valued (e.g., we would not be cutting journals such as *Science*).

A great deal of data was involved to get to the point of proposing the 1,112 journals (table 1). We collected data such as cost for each journal title, if the title was part of a package or if it was paid for singly, usage statistics from the publishers/providers, journal impact factor data and publication and citation patterns by NCSU researchers. We also collated information on whether or not a journal was also available via another access point besides our subscription (i.e., aggregator databases such as Academic Search Premiere).

Table 1: Data elements used at the NCSU Libraries for our serials review exercise

Data	Source
Full list of subscribed serial titles	ILS, Homegrown ERM (E-Matrix)
Cost information for each title	ILS, Homegrown ERM (E-Matrix), Ulrichs, Publisher website
Publisher of each title	ILS, Homegrown ERM (E-Matrix) , Ulrichs
Vendor supplied online usage data (where available)	Vendors use reports (i.e., full-text downloads of journal articles)
Impact factor (where available)	Journal citation reports (Thomson ISI)
Publication and citation data related to each title (where available)	Local Journal Utilization reports (Thomson Reuters)
Alternative access via databases that aggregate full-text journal content	Homegrown ERM (E-Matrix)
Campus feedback on each title	Homegrown web form application

Gathering Feedback

Feedback from campus was central to the process. We did not want to make decisions on titles without extensive campus-wide input. We also wanted a relatively “low-barrier to entry” approach to providing feedback. From previous reviews we knew that posting long, non-sorting lists on the Libraries’ website and inviting email or in-person feedback did not garner as many responses as were necessary. For this review we decided to create a web form where users could easily record and submit their responses and the included tools to help users filter and manage the lists. We worked with our Digital Library Initiatives department to create this form (Figure 1). We presented key data points to enable campus to make decisions on keeping or canceling a title but we did not overwhelm them with all the data points we had available to us.

Figure 1: Screenshot of collections review webform used for the NCSU Libraries 2009 serials review.

Collections and Journals Review 2009

Welcome, Annette Day!
[How to Use This Tool](#)
[Collections Review Home Page](#)

Name:
Comments:
Email:
Logout
Affiliation:
Admin Tools
(Access restricted)
Status:

Customize your list of titles:
[Full List](#)
[My Rankings](#)

1112 Titles

Title ? ▲▼	Publisher ? ▲▼	Call Number ? ▲▼	Subject ? ▲▼	Unit Price ? ▲▼	2007 Downloads ? ▲▼	2008 Downloads ? ▲▼	2007 Impact Factor ? ▲▼	Publications by IICSU authors ? ▲▼	Cited by IICSU authors ? ▲▼	Ranking
A Journal of Church and State	J.M. Dawson Studies in Church and State	BV630 .A1 J6	Humanities and Social Sciences	\$39.00	0	0	0.00	0	1	None ▼
AAPG bulletin	American Association of Petroleum Geologists	TN860 .A3	Agriculture and Life Sciences	\$395.00	0	0	1.27	0	10	None Must keep Keep if possible Can cancel
Abstracts with programs	Geological Society of America	QE1 .G35	Natural Resources	\$162.00	0	0	0.00	0	0	None ▼
Academe	American Association of University Professors	LB2301 .A3	Education	\$68.00	0	0	0.00	0	1	None ▼
Academy of Management perspectives	Academy of Management	HD28 .A244	Management	\$120.00	6	0	0.59	0	0	None ▼

The web form required authentication using campus credentials which allowed us to pre-populate the form with data such as name, email, affiliation and status. The form could be sorted by any of the data points available and filtered by broad subject groupings (Figure 2). Users could also search for specific titles. We offered three possible rankings for a title and suggested that users ignore titles that were not relevant for their field/interests. The rankings were:

- Must keep
- Keep if possible
- Can Cancel

The form also provided a comment box so that a user could communicate additional thoughts on the titles and review process.

Figure 2: Webform demonstrating subject grouping

The screenshot shows a webform interface for subject grouping. At the top, there is a dropdown menu labeled "Filter Full List by Subject" with a list of subjects including Agriculture and Life Sciences, Design, Education, Engineering, Humanities and Social Sciences, Management, Natural Resources, Physical and Mathematical Sciences, Textiles, and Veterinary Medicine. To the right of the dropdown is a search bar labeled "Search Full List by Title". Further right, there are links for "Customize your list of titles: Full List My Rankings". Below these elements is a table with the following columns: "Subject?", "Unit Price?", "2007 Downloads?", "2008 Downloads?", "2007 Impact Factor?", "Publications by HCSU authors?", "Cited by HCSU authors?", and "Ranking". The table contains several rows of data, including titles like "A.J.P.G. bulletin", "Abstracts with programs", "Academe", "Academy of Management perspectives", "Accounting & tax index", and "Accounting and business research". Each row has a "Ranking" dropdown menu set to "None".

	Subject?	Unit Price?	2007 Downloads?	2008 Downloads?	2007 Impact Factor?	Publications by HCSU authors?	Cited by HCSU authors?	Ranking
1 J6	Humanities and Social Sciences	\$39.00	0	0	0.00	0	1	None
A.J.P.G. bulletin	American Association of Petroleum Geologists	\$395.00	0	0	1.27	0	10	None
Abstracts with programs	Geological Society of America	\$162.00	0	0	0.00	0	0	None
Academe	American Association of University Professors	\$68.00	0	0	0.00	0	1	None
Academy of Management perspectives	Academy of Management	\$120.00	6	0	0.59	0	0	None
Accounting & tax index	UML	\$659.85	0	0	0.00	0	0	None
Accounting and business research	Institute of Chartered Accountants in England and Wales	\$387.38	0	0	0.00	0	0	None

The form could be saved so that a user could come back and complete his/her rankings over a period of time before submitting them. Once submitted they were captured in a large comma-delimited spreadsheet for review and manipulation by collection management librarians.

Manipulating the feedback data

We were surprised at how much feedback we received via the form. 1,365 users logged into the webform and of those, 700 submitted feedback. We received a total of 12,710 journal title rankings on the 1,112 titles on the form. In most instances we saw users only rank titles that were relevant to their subject interests. But in some cases we had patrons who ranked every title or ranked a broad spectrum of titles clearly outside their scope.

This presented us with a lot of data to work with and we needed to find the best way to process it in a uniform way for decision-making. We used the following metrics/methods:

- Weighted approach: to minimize the impact of ranking journals outside one's discipline/research areas
- Cost per use
- Combined data points metric
- Collection management librarians' knowledge of campus needs/expectations

Weighted Ranking – Broad approach using college affiliation and journal subject from web form

We weighted the rankings by the community of users who provided feedback based on how closely their research and teaching subject areas matched the journal subject areas. This approach was to help minimize the tendency of users to want to cancel journals that were not relevant to their research and teaching (e.g., a biology researcher may have issued a suggestion to cancel all history journals). The following is an example of a weighting scheme that was used:

- Weight of 1.0 for direct associations between users and journal subject areas
- Weight of 0.8 for close associations between users and journal subject areas
- Weight of 0.5 for partial associations between users and journal subject areas
- Weight of 0.1 for tangential or unrelated associations between users and journal subject areas

We assigned points for each ranking of “must keep,” “keep if possible,” and “can cancel.” This point system allowed us to prioritize those journals that were regarded by our users as titles that should be retained. The scheme for assigning points to the ranking system is as follows:

- Must keep rank = 10 points
- Keep if possible = 5 points
- Can cancel = 1 point

The ranking points were then multiplied by the association weight factors and the total number of rankings, then summed for each journal title. The higher results brought to the surface the titles for which the more closely associated users ranked journals that they deemed were necessary to keep in the collection. See Figure 3 for an example of how the metric was calculated on feedback received for the title *Astronomy Letters*.

Figure 3: *Astronomy Letters* example of weighted ranking

Ranking	Patron's Department	Patron's College	% Match	Weighted Ranking	Weighted Ranking x % Match	(Weighted Ranking x % Match) x Total # Rankings	Sum ((Weighted Ranking x % Match) x Total # Rankings)
Can Cancel	Entomology	Agriculture and Life Sciences	0.5	1	0.5	7.5	621
Can Cancel	Agricultural & Life Sciences	Agriculture and Life Sciences	0.5	1	0.5	7.5	621
Can Cancel	Agricultural & Life Sciences	Agriculture and Life Sciences	0.5	1	0.5	7.5	621
Can Cancel	Poultry Science	Agriculture and Life Sciences	0.5	1	0.5	7.5	621
Can Cancel	Engineering	Engineering	0.8	1	0.8	12	621
Can Cancel	Humanities & Social Sciences	Humanities & Social Sciences	0.1	1	0.1	1.5	621
Can Cancel	Physical & Mathematical Sciences	Physical & Mathematical Sciences	1	1	1	15	621
Can Cancel	Chemistry	Physical and Mathematical Sciences	1	1	1	15	621
Can Cancel	Chemistry	Physical and Mathematical Sciences	1	1	1	15	621
Can Cancel	Mathematics	Physical and Mathematical Sciences	1	1	1	15	621
keep if possible	Engineering	Engineering	0.8	5	4	60	621
keep if possible	Veterinary Medicine	Veterinary Medicine	0.5	5	2.5	37.5	621
Must keep	Electrical Engineering	Engineering	0.8	10	8	120	621
Must keep	Physics	Physical & Mathematical Sciences	1	10	10	150	621
Must keep	Physics	Physical & Mathematical Sciences	1	10	10	150	621

Processing the Feedback – Other metrics used in serials review process

In addition to the feedback and rankings provided by the NCSU campus community of users, we also factored in other metrics such as cost per use of journal titles and journal impact factors, as well as citation and publication patterns by NCSU researchers in the journals under review.

Cost per use: When the cost data and the usage data were both available, the cost/use metric made a very compelling argument for keeping or cancelling a journal.

Where available, other data points such as journal impact factors and citation and publication patterns were used in a variety of formulas to help in our iterative decision-making process. Values of journal impact factors and citation and publication data were derived from Thomson ISI/Thomson Reuters databases (Journal Citation Reports for the impact factor data and Local Journal Utilization Reports for the publication and citation data). In order to take into account these metrics with the feedback from campus and the usage data, various formulae were designed to give more weight to data points we valued highly and reflected a journal's relevance to the NCSU community.

An example of the kind of method that combined this data is as follows:

- Sum of the following:
 - Average of 2 most recent years of use data
 - Number of citations by NCSU researchers to the journal
 - (2 x Number of publications by NCSU researchers in the journal) x (impact factor +1)

See Figure 4 for an example of the various metrics applied to a cross-section of titles in the serials review. In figure 4, “data metric” refers to the calculation described in the preceding paragraph. Figure 4 shows that the different metrics can be inconclusive. For example the title *Materials Science and Technology* had very good feedback from campus as evidenced by its high weighted ranking, but it scored low on the data metric due to its low impact factor and use data. We treated situations like these (where the data was conflicting) as opportunities to follow-up with users who provided feedback to further understand the motivating factors for keeping a low use journal.

Figure 4: All calculated metrics used in the NCSU Libraries serials review process.

Journal Title	Price	2007 Use	2008 Use	Impact Factor	LJUR Pubs	LJUR Citations	Data Metric	Cost per Use	Weighted Ranking
Environmental Progress	\$486.00	64	67	1	0	11	24.62	\$7	165.2
Robotics and autonomous systems	\$1,841.00	107	200	0.633	3	12	34.41	\$12	536
Computational intelligence	\$858.00	23	76	1.972	2	4	26.72	\$17	536
Sensor Review	\$2,972.00	156	84				2.40	\$25	109.9
Journal of environmental science and health - part A	\$3,886.00	99	164	0.967	1	36	79.92	\$30	625.3
Information Processing Letters	\$2,238.00	42	83	0.66	2	10	25.32	\$36	378.9
Materials Science and Technology	\$2,180.00	57	55	0.713	0	0	1.92	\$39	1086.4
Separation science and technology	\$8,678.00	56	172	1.048	0	28	62.01	\$76	284.9
Circuits, Systems, and Signal Processing	\$1,407.00	12	18	0.456	0	2	3.35	\$94	369.9
Distributed and Parallel Databases	\$927.00	6	11	0.771	0	1	2.07	\$109	71.4
Applied Artificial Intelligence	\$1,485.00	15	12	0.753	1	8	18.00	\$110	347.4
Plastics, rubber and composites	\$1,489.00	11	10	0.431			0.30	\$142	80.4
Acta Informatica	\$1,219.00	4	7	0.8	1	7	16.40	\$222	1413.3
Cybernetics and Systems Analysis	\$3,368.00	8	16				0.24	\$281	50.5
International Journal of Satellite Communications and Networking	\$412.00	0	2	0.284			0.03	\$412	254.8
Chemical Engineering Research and Design	\$1,692.00	0	2	0.837	2	22	47.80	\$1,692	151.2

The variety of methods that we used in our serials review process was both instructive and contradictory. But the combination of journal data and campus feedback enabled us to get a summarized view of use (in the broad sense, encompassing full-text downloads as well as publication and citation patterns) and value relative to the campus community. Our approach

enabled us to handle large amounts of data in a consistent way and provided us with broad markers of high and low rankings. Even with weighted methods, we still deferred to strong feedback from campus.

Issues/Challenges

While this may sound surprising to anyone who doesn't work in collections or acquisitions, getting a complete list of our subscriptions with their associated formats and costs was the most challenging part of the project. Today's ILS' aren't set up to reflect how we have to purchase journals today, they don't cope well with packages, bundles, or any kind of dependencies. As a result getting the data we needed was a struggle.

Data points were missing for some titles, particularly those in humanities and arts areas. Counter data was not available for all of our online journals and many titles did not have impact factors or LJUR data. As a result, we could not calculate and compare data metrics across the whole list of titles considered for potential cancellation.

We relied heavily on Microsoft Excel and Access to process the data. This did not require a particularly advanced skill level to use these tools; however, a familiarity and confidence with these tools was needed.

The webform which was so useful in gathering feedback was developed by our Digital Libraries Initiative Department. The development of this webform and the script to collate the data required a skillset that was outside of the Collection Management department's scope.

COLLECTION VIEWS DATABASE

Project Overview

We conceived and built the Collection Views Tool to help us answer the following questions:

- How do the NCSU Libraries' expenditures on resources support the research and teaching needs of diverse colleges and departments at NCSU?
- What data exist that might help us understand how our resource expenditures look in terms of the departments we serve?

We wanted to analyze how our collection supports our campus in a way that brought more insights than just assessing how much money we spend in a specific subject area on a specific format type. For example, having data on how much we spend on chemistry books and journals does not present the whole picture on how we support the Chemistry department at NC State. We know that resources from other subject areas are heavily used by chemists as well. In collaboration with a tech-savvy librarian (Cory Lown), we built the Collection Views tool which allows us to see the full interdisciplinary picture of our collections expenditures and how those expenditures support the colleges and departments at NCSU.

Data

For this project we needed to collect two sets of data: data on library expenditures and data about our institution

Library Data

We collected library expenditures data on all purchased resources (both electronic and print). Portions of the library budget devoted to salaries and other overhead were not included. Most of the expenditures data were collected by running reports within our Integrated Library System (ILS). Data on expenditures for materials acquired through our approval plan were acquired from our vendor, YBP's reporting system Gobi. In addition to expenditures, we also gathered data on the quantities of approval plan and firm order monographs purchased.

In order to match resources to campus departments, we leveraged our subject fund codes (129 in all). The codes are used for collection planning, allowing us to allocate and track funds across different subject areas. Some fund codes are very broad (e.g., General technology -- GTEC) and some are narrow (e.g., Entomology - ENTO). Whenever an item is acquired, a fund code is assigned that best matches the subject area of the item. The subject fund codes were developed to express the research and academic strengths and the cross-disciplinary nature of research and teaching at NCSU.

NCSU Data

At NCSU, the Office of University Planning and Analysis (UPA) collects and manages statistics about students and employees of the university. Basic reports are available on UPA's website but we needed more detail than what was offered; we needed to capture a data-centric view of NCSU. We worked directly with the UPA staff to get headcounts for each department across a number of factors: counts of tenure track and non-tenure track faculty, enrollment (broken down by master, bachelor, doctorate, and associate degree programs), as well as degrees awarded (also divided by master, bachelor, doctorate, and associate levels).

Because research grant income is a major source of research funding at NCSU we also collected total research grants awarded in dollars for each academic unit. We were able to harvest this data from the website of NCSU's Sponsored Programs & Regulatory Compliance Services, which tracks research grant proposals and awards.

Connecting the Data

We wanted to connect the library expenditures data and the campus department data in order to look at these sets of data side by side and calculate metrics such as library expenditure per faculty member in a given department or college. To do this we had to create a mapping between each department and the subject fund codes. Creating an accurate mapping was central to the utility of the tool but it was not easy. The process involved extensive conversations with the collection management librarians in order to make the most robust mapping between subject fund codes and campus departments. We are still refining our method for mapping departments with fund codes.

To create the mapping we established a couple of guidelines to simplify the process. (1) A fund code subject can be applied to more than one department. For instance, some fund codes are quite broad and apply to many departments such as the GSCI (general science) fund. In our mapping any department on campus that had a science element had the GSCI fund code mapped to it. However, it is possible that only a portion of the materials purchased within the fund are truly relevant. (2) The expenditure amount associated with the subject fund code

can chose departments to compare, specific fiscal years and select whether to include all funds in the mapping or narrow down to just specific fund codes (see Figure 6).

Figure 6: Collection Views Data Portal

[Department Overview](#)
[Departments](#)
[Colleges](#)
[Compare Across Categories](#)
[Fund Code Tree Map](#)
[Fund Code Mapping](#)
[Download Data](#)

Compare NCSU Departments

Average Department
2008-2009

Physics
2008-2009

☒ Level 1 (Broad Funds)
☒ Level 2 (General Funds)
☒ Level 3 (Reference Funds)
☒ Level 4 (Branch Funds)
☒ Level 5 (Subject Specific Funds)

	Average Department (2008-2009)	Physics (2008-2009)
Faculty Count	20	32
Enrollment	412	242
PhD's Awarded	8	13
Grant Income	\$2,196,572	\$6,282,023
Total Expenditure	\$1,015,411	\$1,907,014
Monograph Expenditure	\$221,596	\$287,155
Serial Expenditure	\$580,280	\$1,462,982
Database Expenditure	\$213,535	\$156,877

The portal generates graphical output to illustrate comparisons between departments or colleges. See Figure 7 for the output generated in a comparison of the Physics department and the average department at NCSU for the fiscal year 2008-2009.

Figure 7: Graphical output comparing the Libraries support for the Physics department compared to the Average department at NCSU.

Uses of the Collection Views Tool

The broad purpose of this project was to help collection management librarians evaluate and think about how library purchasing power is distributed across different subject areas. The data allows us to examine the relationship between our spending on resources and the departments the library serves, and can help us test our assumptions about the strengths of the university and what we know about the work of colleges and departments on campus.

The outputs from the tool and the kind of comparisons it displays make for good supporting materials for library and campus discussions on budget allocation. Using graphs and charts can simplify otherwise confusing data and allow us to illustrate the data for presentations and impact.

Challenges/Issues

A limitation that needs to be kept in mind when looking at the outputs of this tool is that simple measures across a few characteristics of academic departments cannot tell us everything we need know about differences in research and teaching needs across departments. The data we are utilizing in this tool cannot account for differences in the way that departments make use of library materials. Some collecting areas may be important because they are historical strengths of our collection, even if they do not provide immediate support to particular departments.

Another limitation is that absolutely everything is dependent on the mapping of fund codes to campus departments. That mapping impacts the outputs. For future use we are considering adding weighting to the fund codes to allow for more precise mappings. Gathering the data on the NCSU campus departments in a timely fashion will be difficult as we move forward. That data is from external sources to the library and we have no control over when it will be available to us.

Finally, since we needed SQL programming skills which are outside the scope of the Collection Management department, we leveraged the expertise of our Digital Library Initiatives department to help us create this tool.

RETURN ON INVESTMENT OF JOURNAL BACKFILES – ANALYSIS OF THE COST/USE FOR JOURNAL BACKFILES/ARCHIVES

Project Overview

Using one-time funds supplied by campus, the NCSU Libraries has been purchasing online journal archives (or "backfiles") for the past several years and this has been a substantial financial investment. We wanted to be able to demonstrate the impact of these journal archives not only in terms of usage, but also in terms of fiscal effectiveness. Based on the usage alone, the impact of these journal archives has been tremendous. In 2008 and 2009 alone, the number of downloads of articles in these journal archives has surpassed 350,000 downloads/year.

We took an approach that is a non-traditional twist on the return on investment (ROI) metric. We measured the cumulative cost of the backfile over time compared to the cumulative use of the backfile over the same time period. The metric should get smaller over time as we are looking at the cost of the investment getting closer to paying for itself. This is indeed what we have been seeing. Our historical return on investment (ROI) for the journal archives purchased by the NCSU Libraries has gone from \$6.46 to \$1.07 (since 2004).

Data

The data was derived from a combination of sources including publisher/provider websites (specifically, the administrative modules that provided access to usage statistics), our local ILS and files where cost data was maintained.

We collected usage statistics (full-text article downloads) via journal usage reports from publishers and providers including Elsevier, JSTOR, Wiley, Oxford University Press, Taylor & Francis, Royal Society of Chemistry, Springer, Sage, Nature Publishing Group and Blackwell (prior to their acquisition by Wiley). We also collated cost data for each of the backfile collections that we purchased since 2003. The cost data included both one-time payments and annual maintenance fees.

Calculations

Using Microsoft Excel, we plotted cost/use for each year where the data was available. In many instances we could not acquire usage data back to the year of initial purchase. We carried the one-time cost through each year of ownership and divided that by cumulative use. Where one-time payments existed, we applied the one-time payment to the first year of reported use. For example, for the Royal Society of Chemistry (RSC) Archive, we purchased it in late 2003 and could not get reported usage until 2005 (see Figure 8). The initial ROI was based on the use in 2005 and included the one-time purchase cost as well as the annual maintenance fee for 2004 and 2005 (the ROI = \$13.48 cost/use). As we continued to pay maintenance fees and track usage over the following five years (this data was reported through 2009 for this project), the

ROI dropped from \$13.48 in 2005 to \$2.58 in 2009 since this is a cumulative ROI metric (see Figure 9).

Figure 8: RSC Archive source data

One Time Price	Use Data	Annual Fee	Use Data	Annual Fee	Use Data	Annual Fee	Use Data	Annual Fee	Use Data	Annual Fee	Use Data	Annual Fee	Use Data
2003	2003	2004	2004	2005	2005	2006	2006	2007	2007	2008	2008	2009	2009
\$40,800		\$800		\$800	3145	\$800	3253	\$800	3918	\$1,095	4697	\$1,095	2871

Figure 9: RSC Archive ROI over time

We also summarized the cost/use across all backfile collections and reported an overall cost/use ROI for all our measurable backfiles for each year from 2003 through 2009 (see Figure 10).

Figure 10: ROI for all backfiles

Impact of data

This data was provided to our library administration to use in communications with various administrative and oversight groups across campus. Our goal was to justify our spending of valuable state resources and illustrate the value of the resources the NCSU Libraries purchased and supported. We also wanted to describe how online journal archives saves the Libraries valuable space for seating and innovative work/study/play space, while also enhancing ease of access from virtually any location, noting that portability and discoverability were important needs for NCSU faculty, staff and students. The acquisition of journal archives ensures that our journal collections are comprehensive, spanning access from volume 1 through to present (where we also subscribe to the current issues online).

In terms of internal use, this kind of analysis supports our ability to monitor performance and guide our future approaches in dealing with this kind of content. Determining which online journal archive collections have provided the best initial ROI and what is the impact of those that are drawing on our limited budgets year after year with their annual maintenance fees is key to these approaches.

Challenges/Issues

This project was not without its limitations. Usage data going back to the year of initial investment (back to 2002 in some cases) were not always available. Usage data for some backfiles are partial (Elsevier, for example, does not provide full year data). Usage data for backfile content for some publishers was indistinguishable from usage of current content.

FINAL THOUGHTS

For NCSU, data is a valuable tool. It has been extremely helpful and instructive for us to have data on hand to aid our collections decisions and allow us to articulate clearly some of the decisions we have made. Data is a powerful collection management tool when used in an informed way, but it should not be the only factor in your decision making.

Moving forward at NCSU we plan to continue incorporating data into our collections decisions. We are working on increasing all collections librarians' data skill sets both in terms of tools, focusing on common desktop applications (e.g., Excel and Access) which have proved to be core tools for us, and also working to increase competencies and comfort levels in manipulating and interpreting the data.

We plan to continue to work collaboratively with our colleagues across the NCSU Libraries to leverage skill sets outside of those in the Collection Management department. Not only do these collaborations result in valuable tools to aid collection decision-making, but they also highlight the complexities of collections issues to colleagues across the institution.