Building a Comprehensive Serials Decision Database at Virginia Tech

Paul Metz and John Cosgriff

Although for many years academic libraries have relied on data on cost, library use, or citations to inform collection development decisions respecting serials, they have not fully exploited the possibilities for compiling numerous measures into comprehensive databases for decision support. The authors discuss the procedures used and the advantages realized from an effort to build such a resource at Virginia Polytechnic Institute and State University (Virginia Tech), where the available data included the results of a zero-based faculty survey of serials needs.

Both the growing reliance of internal library operations on automated systems and the migration of publications to electronic format allow new opportunities for the capture and manipulation of management data that libraries can use to optimize the investments they make for their patrons. The use of circulation and reshelving data for this purpose is well established and routine. Spurred in part by the International Coalition of Library Consortia (ICOLC), vendors of online library resources are providing increasingly useful data on searches, sessions, and connect time, making it possible for libraries to compare the marginal return across a range of information resources.

Although libraries have made good use of isolated snippets of management information, few have fully exploited the potential to assemble comprehensive decision-support databases bringing together all known elements about given resources. Consequently, in weighing the value of a resource based on known circulation data, librarians might find themselves wondering about in-house use, for example, and hesitating to make decisions on the basis of limited information. Measures of electronic usage might provide another limited view. What is needed is a medium that brings together everything a library knows or can learn about its resources. This report describes Virginia Tech's efforts to build such a resource for decision support related to serials. It describes the data elements that were obtained and the means by which they were assembled into a total package making possible the comparison of complementary data points in the analysis of serials, the largest single component of the libraries' budget.

Virginia Tech's efforts were not the first to bring together multiple data elements bearing on serial titles, but they were the most ambitious to date. A useful example

Paul Metz is the Director of Collection Management and College-Based Services and John Cosgriff is the College Librarian for Engineering in the University Libraries at Virginia Polytechnic Institute and State University; e-mail: pmetz@vt.edu and cosgriff@vt.edu, respectively. *The authors wish to acknowledge the work of Linda M. Doss, Lori Lee, and Milko Maykowskyj in preparing the data for this study.*

of the effort to obtain a cumulative value through the use of multiple measures is Janet Hughes's use of both the Institute for Science Information's (ISI's) *Journal Citation Reports* and data on the citing and publishing practices of Penn State faculty in the life sciences to determine a core list of serials in molecular and cellular biology.[1]

Virginia Tech's efforts were not the first to bring together multiple data elements bearing on serial titles, but they were the most ambitious to date.

Serials Information at Virginia Tech

In 1997, Virginia Tech's vice president for information systems appointed a Task Force on the Digital Library to study the means by which the libraries could maximize scarce resources available for serials, while accelerating their emerging reliance on digital resources readily accessible to remote users. The task force included several library representatives and a number of well-respected teaching faculty from all areas of the university. The impetus for the task force had two sources, one a problem and the other an opportunity. The problem was that the combination of runaway serials inflation and relatively flat funding had led to multiple rounds of serial cancellation, in which more than 4,500 titles had been canceled in the 1990s alone. The methodology used by Virginia Tech to cancel its serials has been reported elsewhere.[2] Cancellation decisions were made in close communication with the faculty but initially were triggered by examination of all data available at the time, including price, overall citation rankings from *Journal Citation Reports*, and periodic samplings of the reshelving of current periodical titles.

The opportunity was the emerging marketplace of digital resources—the possibilities of which both Virginia Tech as an individual institution and the Virtual Library of Virginia (VIVA), of which Tech was an active member and beneficiary, already had been aggressively pursuing—and the growing availability of in-house data sources.

Among the expectations with which the task force was charged was that of making a credible estimate of the size and cost of maintaining core serials sufficient to support the university's main missions in research and instruction. The faculty on the task force suggested that the only way to accomplish this would be to ask the entire faculty to identify the serial titles they found essential to their work.

The librarians on the task force supported this recommendation, though with some reservations. Having involved the faculty in successive rounds of serial cancellation, the librarians had hoped to let a year pass without imposing on faculty time again. Moreover, the librarians were concerned that a broad solicitation of faculty input on serials might be construed as a hidden mechanism for canceling them. In addition, there was concern about the disparities and errors that inevitably would occur as faculty members tried to remember serial titles and report them in their own ways. On the other hand, of course, faculty opinion about which serials were most needed would be invaluable, and it was expected that the origin of the request in a task force representative of faculty would promote a high level of faculty participation.

After the task force report had been accepted, an internal library work group was established to determine the best means of soliciting faculty input. It was quickly decided that a Web-based questionnaire, based loosely on the libraries' past success in using the Web to solicit faculty responses to serials nominated for possible cancellation, would be the ideal medium for the solicitation. Various means of incorporating an automated mechanism for assisting respondents and thus yielding greater title authority, such as the automatic interposition of URLs for Ulrich's or Publist information in the online questionnaire, were studied but rejected as likely to be so cumbersome as to greatly depress the response rate. Instead, a hot link to Publist

FIGURE 1
Factory Serials Survey

Assessment Form
At $4 million per year, the Libraries' serials expenditures account for nearly a full percent of the University's budget. A number of your colleagues have suggested that it is time to give each faculty member, including those hired in the last few years, the opportunity to consider the entire spectrum of their journal and serial needs so that they can very directly **tell us what is wanted.** We agree! We very much want to understand your specific needs so we can include this information as part of the budget request and planning cycle for 1999/2000.

In the form below, please furnish at least the titles for journals, "advances in" series, indexes, databases, or other serial publications useful to your teaching and research to which you believe the University Libraries should subscribe. We are equally interested in titles currently held and those not now in our collections. It is not necessary to include obvious core titles such as *Nature, Daedalus, or The Harvard Educational Review.*

While only each title and identifying information about yourself are required fields, we would appreciate any further information, including comments.

If you are unsure of a title, you may verify it by calling the main library reference desk at 231-9232 or try the PubList. For general questions about this project, please consult the FAQ.
Use the tab key to go to the next field. The form submit button is at the bottom of the page. **Thank you for participating in this important project.**

Lastname	**Firstname**
Department	

If you need more than **10**, select how many you need below first. Please request the largest number of title lines you could possibly need. Blank records are ignored.
20 30 40 50 60

Item	**Title**	**Publisher (*Optional*)**	**ISSN (*Optional*)**	**Comments (*Optional*)**
1.				

was provided for respondents to use, at their option, to verify serial titles.

The Web site was put up early in the spring of 1999. Multiple solicitations in a variety of media, including e-mail, articles in the campus newspaper, and meetings with faculty, were used to encourage response. The libraries stressed that although future serial cancellation projects were likely and would benefit from the solicited data, the immediate purpose of the survey was to identify a true core of most-needed serial titles, including any to which the libraries might not have a current subscription. The text of the Web site to which survey participants were directed is reproduced in figure 1. All publicity efforts in other media closely followed the wording used in the site.

The FAQ sheet to which the survey page was hot-linked had been created in response to both actual and anticipated faculty questions. Paper copies were made available at reference sites and given to all college librarians. It is reproduced in figure 2.

The Web site was maintained throughout the spring semester. With multiple reminders and much urging, the database grew to match all but the most optimistic projections. More than four hundred faculty members responded, casting more than 9,000 votes on behalf of more than 4,000 distinct serial titles. To some degree, the libraries' fears were realized. It was difficult to disabuse some faculty of the most cynical interpretations of the project. And as expected, the challenge of cleaning up free-text nominees was considerable. Idiosyncratic abbreviations, word omissions or inversions, and inconsistency of entry, especially title versus corporate author, posed a challenge that took hundreds of hours to (largely) overcome. The instructions to omit references to highly obvious titles such as *Science*, which had been put in at the faculty task force's insistence and despite the librarians' reservations about the subjectivity and error this might introduce, apparently were not followed because *Science* and similar titles received many votes.

After the survey was completed, the libraries had to decide how best to use the data. It was a great advantage to know how many faculty members considered a given title critical to their work, but this was only a single data point. To make fully informed and cost-effective decisions about titles, other measures of demand and interest would be required, as would the key element of price. This realization led to the decision to assemble all other possible data elements bearing on library serials, even if doing so added months to the project.

Gathering and Assembling Serials Data

The first task was to bring together the faculty vote data and to clean them up. The original data first were moved from the Web survey repository to an Excel spreadsheet. Four volunteers worked to expand abbreviations, correct errors, and move generic titles such as "Journal" or "Proceedings" to corporate author entry, wherever possible. The many individual lines built into the Web-based survey were collated into single lines so that, for example, twenty-six lines representing votes for *Science* were reduced to a single line with the value of twenty-six in the "votes" column. Several departments (management, horticulture, finance, theater, physics, building construction) or groups of faculty within a department (the mathematics education faculty) had chosen to submit lists representing their collective priorities rather than to vote as individuals. In all but two cases, these were simply binary, in that a title either was or was not on the critical title list. Building construction and physics used ordinal ranking scales. A column was devoted to expressing these values for each department or group that had chosen to vote in this manner.

If the premise for collating the vote data with other data sources was that the whole would be greater than the sum of the parts, it followed that the whole would be greatest if the number of relevant parts could be maximized. The libraries were fortu-

FIGURE 2
Serials Needs Assessment FAQ Sheet

1. Why are the Libraries asking faculty what serials they need?
   - Because we have never before systematically gathered this information. A limited number of new serials have been acquired each year, nearly all in response to faculty suggestions. Although we believe our decision-making process has been rational, faculty knowledge about the Libraries' procedures and how to submit requests has been haphazard;
   - Because scholarship, science, and the curriculum have greatly changed since the majority of our serial subscriptions were originally placed, while technology has greatly changed the means of access and the nature of the audience we are trying to reach;
   - Because we need to justify our budget in the face of competing university priorities;
   - Because it would be irresponsible not to perform a periodic reality check on an annual investment of $4 million, or nearly a full percent of the university's budget;
   - Because what we learn about serials needs will tell us more generally what the library's "customers" require, suggesting other ways in which we can serve them;
   - Because, within very real financial constraints, we do intend to place subscriptions for some of the more heavily requested titles, and because we will emphasize frequently mentioned titles as we set priorities for what resources to network electronically.

2. Don't you already have these data from all our past serials cancellations?
No. In each cancellation project we have nominated for possible cancellation a fifth or less of our titles. We have comments on those, but no comments on other titles in our collections or on the many titles we don't own.

3. Suppose nobody names a certain title in this survey, will it be canceled?
This is not a cancellation project, and indeed we do not plan to cancel serials this year. We chose this year for the project because it wouldn't be contaminated by worries about serials cancellation. Some titles not mentioned by any faculty respondent would be safe from cancellation because we know they are core titles, because they are covered by our standard indexes or are heavily used by students, or because they support our reference services. Other titles not mentioned by anyone would potentially become candidates in any future cancellation, but they would not actually be canceled without the full review we have instituted each time we've canceled serials.

4. Why are you asking only faculty to participate?
Participation by *anyone* is welcome, but we are actively soliciting faculty input because faculty drive the research agenda and the curriculum. Also faculty will generally be in the community long enough to justify the long-term investment serials subscriptions represent

5. What kinds of publications are in scope?
Anything that isn't a one-time publication. Journals, monographic series, conference proceedings, annual reviews, and "advances in" publications are all in scope. So are indexes, whether print or electronic. So are databases of financial, legal, scientific or other kinds of data, which require periodic updates to remain useful.

6. What about other needs besides serials?
The input form has an optional area for general comments. Use this to talk to us about books and videos, services, book drops, kudos or complaints, whatever you think we should know.

nate that a number of useful components were available locally. The reshelving of current periodicals had been counted on a title basis since June 1998. And the ILLiad system for interlibrary loan (ILL), developed at Virginia Tech, has among its many advantages virtually comprehensive management reports, which include counts of titles borrowed. The counts are updated in real time and for those titles borrowed five or more times are constantly available from the ILLiad Web page. Both of these in-house data sources were mined, and serial titles were moved to the already-present title field and new columns created to hold the counts of reshelving and interlibrary borrowing, respectively. The shelving data included the local call number, which also was added. The original shelving database contained blank values for titles not having been reshelved at all. These were converted to zeroes before moving the data to the final spreadsheet so that the important distinction between "no use" and "no observation" would be maintained.

Although intensive consultation with the faculty will accompany any such future rounds as it has in the past, the libraries expect the decision support available from the data to be invaluable in determining which titles should be suggested for review.

The libraries then turned to external sources for additional data. At the suggestion of one of the original task force members, the libraries contacted CARL Uncover to find out whether management data were available on the number of faculty who had selected table of contents SDI updates for the individual titles available within CARL Uncover Reveal. CARL Uncover Reveal was a service the libraries had been promoting for some time, and it had several hundred Virginia Tech subscribers. These data were available in a simple spreadsheet format and were added, again using the existing title column and adding a column for the new observations.

For more than a year, the libraries had been receiving regular title-level reports on citation retrieval and full-text displays within InfoTrac's Expanded Academic Index. It was not cost-effective to collate the data from multiple months, so April 1999, a high-use month, was selected and these data too were added.

The final indicators of the need for individual titles were obtained when the libraries decided to procure the *Local Journal Utilization Report* (*LJUR*) available from ISI. Unlike the better-known *Journal Citation Reports*, which reports publications, citations, and other derived data points such as half-life or journal impact factor on a global basis, *LJUR* data are tailored to report the citing and publishing practices only of authors identified in the author field as coming from a specified institution. The number of times Virginia Tech faculty had published and the number of times they had cited any of more than 2,500 serial titles in the years 1994–1998 became available through these means and provided the final two columns of data assessing user need.

It is important to concede that not only is each of the data points being considered as an indicator of the demand for or utility of a title flawed, but also that the impression gathered from a simultaneous consideration of all the measures is imperfect. Neither in-house nor electronic usage, nor publishing or citing patterns, nor use of ILL to procure articles from other libraries, nor the sum of all these measures provides a fail-safe indicator of value. Popular, sometimes ephemeral, titles have high usage. Publication numbers do not represent the length or quality of the individual articles published or the size of the readership. It will always be impossible to make highly accurate assessments of cost-benefit ratios for serial titles. Indeed, several studies comparing disparate measures of the demand for or usefulness of serial titles have reported relatively low correspondence among measures such as reshelving and citation counts, reshelving and faculty ratings, or citation counts and interlibrary borrowing.[3] The purpose of

TABL 1

D	t	N Wi	s a	H	Ti e	m	S re
F	9	4	8	A	A	f	38
				A		)	
C	s	1	5	N e			23
	6/	9	6691	P	Week y		738
Virgin	Tech		2601	A	A	y	289
Virgin	Tech	8	2601	J	A	y	977
I	s	1	2	E	A	y	
	9					y	192
I r	A	9	2	0	B t		1 4
I r	A	9	1411		N York Ti s		278

acquiring and assembling the indicated measures was to triangulate toward the truth and reduce error by bringing together complementary kinds of data bearing on different aspects of value and demand.

Table 1 illustrates the various data elements used as criteria indicating the use and value of serial titles, showing the number of cases with data and the highest-ranked two titles for each criterion. The high recognition factor for the titles appearing in the table may serve as an indicator of the "face validity" of the measures.

The Faxon Corporation, the vendor for the great majority of the libraries' serials, was the last external source to contribute data. It provided price, frequency, and a basic class number for the titles to which the libraries had a Faxon current subscription and for which they could provide ISSN (already available in the ILLiad and InfoTrac records).

Uses of the Data

Even after the de-duping of more than 4,000 lines for the individual and faculty votes, the information gathered for the database comprised more than 23,000 lines of Excel data. Various approaches were considered for concatenating all data for each title into a single line, but this proved not to be cost-effective. ISSN, which would have been the only workable hook for the recognition of common titles, was absent from too many of the data sources and could be obtained only through expensive and labor-intensive means.

At this point, it was decided to split the database into two components: a simplified version intended to meet the original goal of identifying and pricing the serials most needed by the Virginia Tech community, and another version that would convey all the detail necessary for title-level decisions relating to either serial cancellation or the acquisition of new titles. Because the provost and other senior officers

TABLE 2
Serials Decision-Support Database (aka the Big Ugly Database)

Source	Frequency	Catalog Title	Call No.	COST 1999	ISSN/ISBN	LC Call No. (Auth)	Lang.	Votes	CARL	ILL	VT Pubs 94-98	VT VT Cites 94-98	Resolved 6/4/98-8/23/99
Vend/Votes/CARL	Semi-monthly	Molecular and cellular biochemistry		$3,672.31	0300-8177	QR	English	4					
ISI 94-98 VT pubs,cites		MOLECULAR AND CELLULAR BIOLOGY			0270-7306						0	102	
Vend/Votes/CARL	Monthly	Molecular and cellular biology	QP187 A1 M65	$460.98	0270-7306	QH	English	8	2				
Current shelving		Molecular and											4
6/4/98-8/23/99		Cellular Endocrinology											
ISI 94-98 VT pubs,cites		MOLECULAR AND CELLULAR ENDOCRINOLOGY			0303-7207						1	16	
Vend/Votes/CARL	Semi-monthly	Molecular and cellular endocrinology		$3,056.64	0303-7207	QP	English	1					
ILL borrows		Molecular and cellular probes								6			
7/1/97 - 8/23/99													
ILL borrows		Molecular and chemical neuropathology /								6			
7/1/97 - 8/23/99													
ISI 94-98 VT pubs,cites		MOLECULAR AND GENERAL GENETICS			0026-8925						1	108	
Vend/Votes/CARL	Semi-monthly	Molecular and general genetics.	QH506 M661	$3,454.32	0026-8925	QH	English	7	2				7
Current shelving		Molecular Biology and Evolution			0737-4038						0	22	
6/4/98-8/23/99													
ISI 94-98 VT pubs,cites		MOLECULAR BIOLOGY AND EVOLUTION		$342.87	0737-4038	QH	English	2					
Vend/Votes/CARL	Monthly	Molecular biology and evolution			1059-1524			2			13	32	
ISI 94-98 VT pubs,cites		MOLECULAR BIOLOGY OF THE CELL											
Current shelving		Molecular Biology of the Cell / ASCB	QH604 C442										11
6/4/98-8/23/99													
Vend/Votes/CARL	Monthly	Molecular biology of the cell		$383.81	1059-1524	QH	English	7	1				

(—> etc. 5 more departments' proxies) IAC Views April 1999; IAC Retrievals April 1999; Math Ed proxy 1=present; Mgt proxy 1=present; Hort proxy 1=present; FIBL proxy 1=present; Thea proxy 1=present; Physics proxy 1=highest, 5, then blank= lowest; Bldg Const proxy 5 respondents, scores range 8(highest) to 60

TABLE 3
Simplified Presentation for Budget Justification

Frequency	Title	Price	ISSN
Semimonthly	Molecular and cellular biochemistry	$3,672	0300-8177
Monthly	Molecular and cellular biology	$461	0270-7306
Semimonthly	Molecular and cellular endocrinolog	$3,057	0303-7207
Semimonthly	Molecular and general genetics	$3,454	0026-8925
Monthly	Molecular biology and evolution	$343	0737-4038
Monthly	Molecular biology of the cell	$384	1059-1524

of the university wanted to know the "bottom line" for a basic serials collection, it was necessary to present these data in a clear and elegant manner free of redundancy.

To assemble a list of titles that would be considered important to the Virginia Tech community, it was necessary to devise specific selection criteria. Ultimately, titles were deemed of significant value and interest if they:

- Received one or more individual or departmental votes
- Were profiled on CARL Reveal by five or more individuals
- Were borrowed twenty or more times on ILL
- Contained ten or more publications by Virginia Tech authors
- Were cited fifty or more times by Virginia Tech authors
- Were reshelved fifty or more times

Views and retrievals on InfoTrac were not considered credible as a selection criterion because it could be argued that titles available through this VIVA-funded means could be canceled locally. Ultimately, 4,563 titles met one or more of these criteria, the great majority qualifying by virtue of faculty votes. For these titles only, every effort was made to clean up the database. All titles meeting the criteria were represented in a new database, with one line for each title. Prices were added for hundreds of titles previously lacking this information. The estimated annual cost of serials in this most-needed list was $3,588,000. This figure was derived by successfully identifying the prices of 85 percent of the titles in the reduced database and applying the per title average of $786 to the remainder.

Tables 2 and 3, each of which represents data for titles beginning "Molecular a" through "Molecular b," illustrate the divergence between the original data set with all its detail and the data presented to the provost and the deans as part of budgetary justification. In table 2, as many as three lines are required to represent all data elements for an individual title, whereas redundant lines have been removed from table 3. Table 2 also includes two titles (*Molecular and Cellular Probes* and *Molecular and Chemical Neuropathology*) that did not qualify for the second list. As has been indicated, the purpose of table 3 is to present the list of most-needed titles in a highly compact manner sufficient to lay out the community's needs and to establish a budgetary requirement. It would be necessary to consult table 2 for the information required to make any title-level collection development decisions or even to identify the reason(s) a title may have qualified for the key title list.

The 23,000 lines of the original data will be used to inform title-level decisions. The opportunity to fully use these data significantly has not yet arisen, in that the libraries' budget is not sufficient for significant numbers of new titles to be entertained. However, a few new titles have been added, based mainly on statistics showing heavy interlibrary borrowing. It is likely that the first significant use of the data will come in the next round of serial cancellation, whenever that may be. Although intensive consultation with the faculty will accompany any such future

rounds as it has in the past, the libraries expect the decision support available from the data to be invaluable in determining which titles should be suggested for review. The libraries probably will cancel somewhat more titles than is essential simply to keep the budget in balance, using the funds from additional cancellations to acquire most-needed new titles identified by the data.

A small example already has shown the potential utility of the data. In the fall of 1999, the libraries received complaints about cancellation of Parts B and C of the title *Transportation Research*. This title is published in parts A through D. The libraries had never owned D, had canceled B and C, and still retained part A. In response to the complaints, the libraries consulted the serials database, which showed no use of any kind for parts C and D. Part B turned out actually to have been cited more than Part A (thirty-three times versus six) and to have carried three Virginia Tech publications against none for Part A. Although both the reshelving and ISI data showed that the libraries had been right not to cancel Part A, they also showed that it had been an error to cancel Part B. Therefore, Part B was reinstated and Part C was not.

Future Implications

Although the artistic elements of collection development will always be important, decisions allowing the best support for a library's community also must rely heavily on a nearly scientific analysis of the best data available. Typically, such analysis has been handicapped by a fragmentation of data elements, making it difficult to examine all relevant information simultaneously. Although it is labor-intensive to do so, complex multidimensional databases can be assembled to facilitate support for the highly expensive and important decisions that libraries make about their collections. Future improvements in the capture of use data, especially through the monitoring of electronic usage, will only increase the possibilities for more sophisticated analysis.

For Virginia Tech, the serials decision-support database already has played a useful role in justifying budgetary needs on a macro level and indicating the most cost-effective decisions on a micro level. Had it been possible to collate the data to a single-line entry for each title, not only greater elegance, but also the ability to experiment with weighted scoring leading to a composite score or to test empirically the correlations among disparate measures of the utility of each title would have been possible. It is hoped that future efforts at Tech or elsewhere will be able to achieve such integration.

Another question with implications for the future is how the libraries will keep their data current. Some data remain useful for a longer period than others. For example, Maurice B. Line has shown that although total citation counts are very stable over time, data on ILL activity at the title level are much more dynamic.[4] In the case of Virginia Tech, it should be relatively straightforward to substitute more current shelving, ILL, and InfoTrac data whenever needed because these are available from either in-house systems or regularly received VIVA reports. Should the libraries begin to measure the reshelving of bound periodicals, as is now being considered, the authors would, of course, leap to the chance to include these data. It should not be difficult to receive new CARL Reveal data once a year or so. On the other hand, Virginia Tech faculty would be unlikely to welcome frequent solicitations of their expressed needs, and the *LJUR* reports from ISI are quite expensive. Based on the value of the data, the premium put on currency, and the cost and bother involved, the libraries will have to decide how frequently to refresh the database.

Notes

1. Janet Hughes, "Use of Faculty Publication Lists and ISI Citation Data to Identify a Core List of Journals with Local Importance," *Library Acquisitions: Practice & Theory* 19 (1995): 403–13.
2. Paul Metz, "Thirteen Steps to Avoiding Bad Luck in a Serials Cancellation Project," *Journal of Academic Librarianship* 18 (May 1992): 76–82.
3. Keith Swigger and Adeline Wilkes, "The Use of Citation Data to Evaluate Serials Subscriptions in an Academic Library," *Serials Review* 17, no. 2 (summer 1991): 41, 46, 52; Marifran Bustion and Jane Treadwell, "Reported Relative Value of Journals versus Use: A Comparison," *College and Research Libraries* 51 (Mar. 1990): 142–51.
4. Maurice B. Line, "Changes in Rank Lists of Serials Over Time: Interlending versus Citation Data," *College & Research Libraries* 46 (Jan. 1985): 77–79.