

Making the most of a 'Big Deal': Building a consortial shared list to reclaim title-by-title eJournal selection for libraries



Statewide California
Electronic Library
Consortium

Jason S. Price, Ph.D.
Life Science Librarian
The Claremont Colleges
jprice@libraries.claremont.edu

Introduction

The Statewide California Electronic Library Consortium began its ScienceDirect contract in late 2001 with a total of 12 institutions signing on after a semester of access to the complete collection. By mid-2004 there were 10 institutions left, and the size of our financial commitment to the deal had grown greatly thanks to the acquisition of Academic Press by Elsevier in 2002. My sciences journal budgets were overspent, and we were seeking ways to control costs within the package. At the same time, we had user requests for some titles we lacked access to, and use statistics showed that we were paying for access to hundreds of titles that were not being used. I lamented the loss of title-by-title control to Elsevier's batch-marketed subject collections, and the Unique Title List they were offering appeared to be a step in the wrong direction. Then came a rare bit of good news: we could build our own lists, as long as they were shared among institutions. Here I provide specific reasons why we felt the subject collections and UTL were inadequate, describe the process of building the shared title lists we assembled as an alternative, and provide an initial evaluation of their effectiveness.

The ScienceDirect 'big deal'

The bad news. Institutional customers agree to be 'locked-in' to a subscription commitment for the life of the deal. They commit to maintain their Elsevier print subscription total plus a fixed annual price increase, often called a price cap (e.g. 0.5x%), presumably because it is lower than the market rate increase.. In 'return', they gain perpetual e-access to these **subscribed titles** for a percentage of their list price, say x%. Although this e-access fee is less than they charge for e-access outside of the deal, it is still a significant additional commitment. Although some contracts include an annual cancellation allowance (0.1 to 0.2x%), it is not large enough to allow libraries to balance their budgets within the ScienceDirect Package (given a fixed annual inflation rate that is at least 2.5 times the cancellation allowance). Thus the big deal requires that libraries forfeit the ability to balance their budgets by cancelling a sufficient number of Elsevier titles when Elsevier's prices go up faster than the library budget.

Putative good news. The benefit that theoretically balances this significant and ever-increasing commitment is low-priced leased access to **unsubscribed titles**. Although

other publishers' Big Deals may not charge separately for this additional access and/or may open their entire collection, Elsevier charges fees which usually limit libraries to a subset of the complete collection.

These collections take one of two forms. The older model is to lease selected pre-packaged subject collections with prices apparently varying from 0.1 to 0.5x% of list. Presently, there are 24 subject collections containing 1788 Elsevier titles. The newer model is to lease access to a Unique Title List (UTL) which is based on the consortium's subscriptions. UTLs consist of a deduplicated list of all titles subscribed to by at least one member of the consortium and are priced similarly to subject collections. Both models have significant drawbacks.

Shortcomings of Subject Collections. Subject collections are the worst of these two evils, for a number of reasons.

- Agricultural and Biological Sciences
- Arts and Humanities
- Biochemistry, Genetics and Molecular Biology
- Business, Management and Accounting
- Chemical Engineering
- Chemistry
- Computer Science
- Decision Sciences
- Earth and Planetary Sciences
- Economics, Econometrics and Finance
- Energy
- Engineering
- Environmental Science
- Immunology and Microbiology
- Materials Science
- Mathematics
- Medicine and Dentistry
- Neuroscience
- Nursing and Health Professions
- Pharmacology, Toxicology and Pharmaceutics
- Physics and Astronomy
- Psychology
- Social Science
- Veterinary Science and Veterinary Medicine

1. **Publisher-fixed subject collections prevent quality-based selection.** By definition, large subject-based collections tend to have a few high-use journals and many low-use journals. Thus libraries that lease subject collections are unintentional accomplices in supporting low quality journals. An unfortunate corollary is that dropping even a relatively low use collection requires the loss of at least a few valuable titles.
2. **Publisher-fixed subject collections force collection gaps.** While they may be suitable for very specialized schools (e.g. Medical, Veterinary or Dental), even a school of natural sciences would need to purchase at least half of the collections to approach comprehensive coverage. This problem often forces decisions as to which departments to support and which to deny.
3. **Publisher-fixed subject collections result in double and triple payment.** Subscribed titles that are duplicated in subject collections are paid for both through the e-access fee and the subject collection fee. Double payment is quite common, because institutions tend to lease collections in subject areas in which they also have a large number subscriptions. Triple 'payment' occurs when these titles are included in more than 1 subject collection. One analysis shows that pairs of subject collections can have up to 50% duplication and the percent of unique titles in any one collection ranges from as low as 5% (!) to a high of just 75% (Ives, 2005). These two common forms of overlap make it meaningless to calculate percent of list price for individual subject collections—accurate values can only be calculated for the complete set of leased titles after the

aggregated subject collection list has been deduplicated internally (i.e., within all leased titles) *and* externally (i.e., against the subscribed title list). Thus discounts are not nearly as large as they appear, because libraries commonly paying two or more times for e-access to the same titles.

4. **Publisher-fixed subject collections are not really fixed!** Titles are added and moved on an annual basis as Elsevier acquires new titles and seeks to optimize marketing of the collections. I am not aware of any case where Elsevier intentionally revoked subsequent access to a title that was part of a collection in the first year of the lease. It is clear, however, that subject collections with the same name purchased in different years will have different title lists and that later versions lack some of the titles contained in earlier versions. Titles are also added to these collections over time. The resulting disarray grows with time as Elsevier attempts to manage 24 subject collections x 5 years and counting. This results in inevitable inconsistencies between library ejournal catalogs and actual availability that do not serve users well. Furthermore, recent changes from
5. **Publisher-fixed subject collections are not priced consistently over time.** For SCELC, at least, recent changes in our invoices make subject collections less attractive. They have recently begun to be priced relative to their total list price rather than equally per collection, though the percentages were adjusted so that there was no affect on the current bottom line. This change suggests that future cancellations or additions will be priced in this way, despite the overlap mentioned above.

Shortcomings of Unique Title Lists. UTLs are an improvement over subject collections. They solve the duplicate payment problem because they are deduplicated by nature, and are discounted based on each institution's subscribed title list. However, their other selling points are dubious at best.

1. **Based on legacy subscription profiles.** Elsevier suggested that gaining access to all of the titles subscribed by other institutions in the consortium would ensure that we were only leasing important titles, and therefore be an improvement over the subject collections. But important to whom? Given that most institutions had far more titles on their leased title (subject collection) list than on their subscribed list, we had little confidence that a list based on subscribed titles would include all of the most valuable. Legacy subscribed titles (so named because they were established far in the past and cannot be changed) become less and less relevant with increasing electronic access and use of leased titles. Our subscription lists are likely to be hazy reflections of the sets of titles that have become most valuable to our users. Put another way, it seemed foolish to give up the high (& low) use titles out of our subject collections in favor of titles that had been chosen by other institutions in the distant past. We knew which titles were most important to our users (at least within our collections)--the ones that had received the greatest online use in recent years.
2. **One size fits none.** UTLs do allow broader coverage, since they do not force choices between disciplines. The flip side, however, is that they are a one-size fits all solution--they do not accommodate specialized schools that might want to focus on particular disciplines. They are also biased toward larger budgets--either you can afford to lease the whole list or you can't. When there are major differences in the size or focus of institutions in the consortium, UTLs favor larger schools in that: a) they can afford to lease a large list, and b) they get a greater discount on the list because they have more overlap between their subscription list and the UTL. Thus Elsevier's UTLs cost smaller schools more than larger schools, lacking size and subject flexibility.

3. Higher priced titles? Priced as a percent of list minus any overlap with an institution's subscribed titles, UTL pricing is indeed more transparent. We also expected, however, that each institution was subscribed to a number of high-priced titles deemed necessary for their programs, or left as a legacy of a more affordable past. Combining these, then, seemed to make it likely that a UTL would contain a preponderance of high-priced titles, especially since these are the very titles that cannot be cancelled under the meager 0.1-0.2x% cancellation allowances. We wanted a chance to eliminate some of these legacy behemoths from our leased title list.

Clearly then, after a brief discussion of the UTL option, our Consortium's ten ScienceDirect subscribers agreed that it was *not* a viable alternative.

Truly good news. When presented with our disinterest, Elsevier's Barbara Kaplan provided the first really good news since the start of our contract: **Elsevier doesn't care which titles are on our 'UTL'**. They were even willing to manage more than one list to allow for differences in size or subject emphasis (even for subsets of our 10 institution consortium). After repeated confirmation that we had understood correctly, we were left with a single question: which titles should be on this *shared title list* or lists? The simple answer: each institution's most used titles. But of course, the devil is in the details...

Building the Shared Title Lists

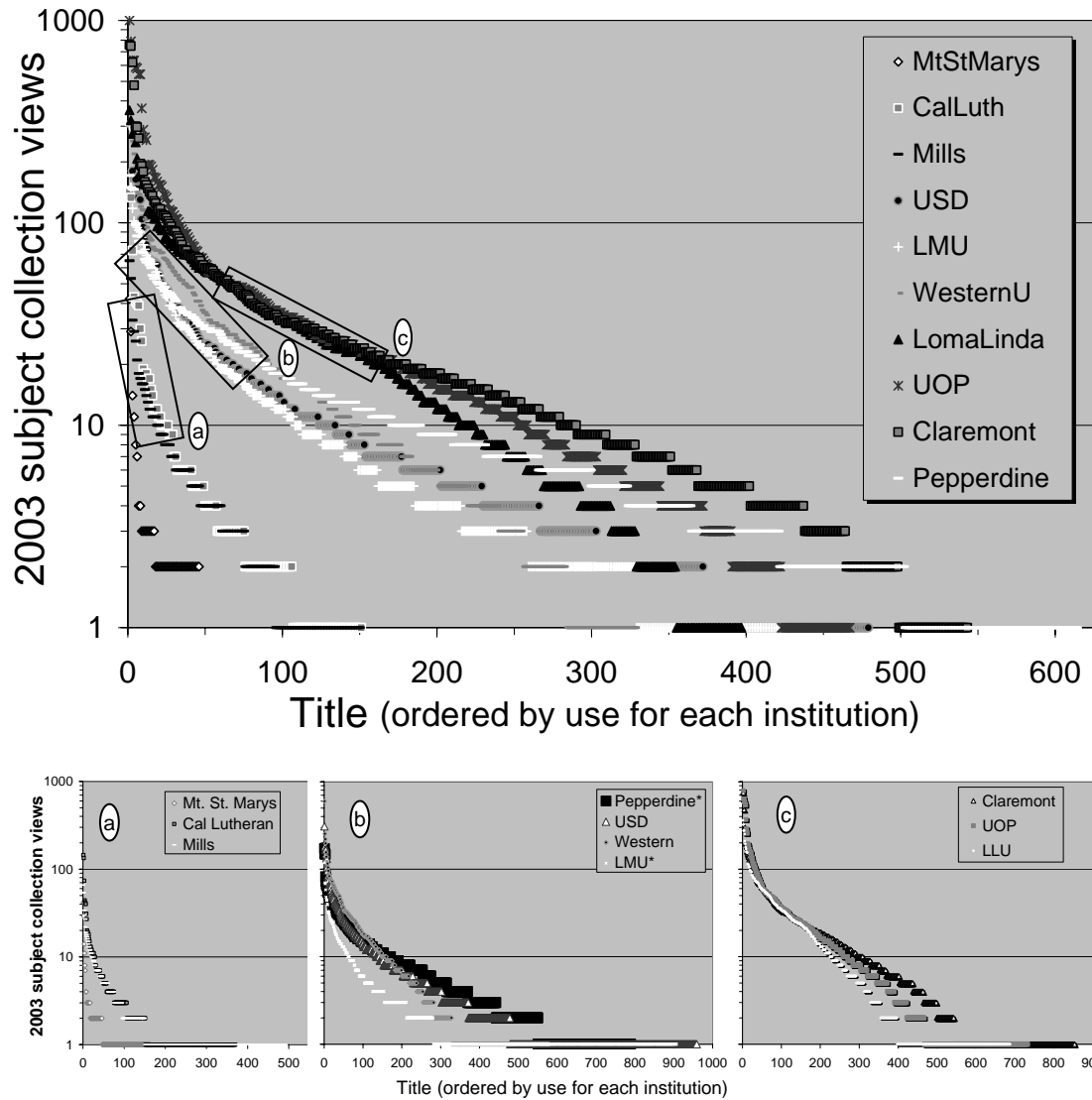
A shared title list (STL) differs from a unique title list in that the Consortium gets to select the titles that are included and can change it as often as once a year. The other significant difference is that a consortium can build more than one list. Otherwise, the terms are identical to unique title lists: access to titles on the list is leased by annual payment of a small percentage (e.g. 0.25x%) of the list price of all the titles on the list that the institution does not subscribe to. Thus every institution that opts for a shared title list has access to the same titles, but pays a different amount for that access.

The goal of this project was to merge each institutions historical subject collection use data to form a shared list or lists that represent the ideal compromise between the subject collections and the UTL, providing access to the most valuable titles at the lowest possible price.

Guiding principles

- 1) **Base the contents of the list on the history of use by the Consortium's users.**
Hereafter, 'use' of an individual article from a title whether a view of HTML or PDF full text will be referred to as a view. The proportion of these articles that are actually used is a subject for another paper—the relationship between views and ILL requests would be of particular interest.
- 2) **Guarantee needs of all institutions are addressed** by including every title that was requested more than a fixed number of times at any institution.
- 3) **Keep the list as small as possible** to cut costs while increasing collection quality.
- 4) **Weed out high price per view titles**, preferentially removing these when possible.
- 5) **Recognize that building this list allows for future flexibility**, allowing the list's shortcomings to be improved title by title in future years.

Figure 1. Patterns of ScienceDirect subject collection use by SCELC institutions. The distributions of use among titles were visualized by plotting the COUNTER 'total number of full text article requests' on a log scale vs. journal title for each institution. The journal title (x) axis was sorted from high to low use separately for each institution. Zero use titles were ignored in the composite graph for clarity. Panels (a) thru (c) include the zero use titles (raw view data was transformed by adding 1 use for every title since log (0) is undefined).



Collecting the data. I collected the COUNTER-compliant statistics for 2003 from all 10 ScienceDirect members with the goal of combining these numbers to select the most used titles from each institution. Most were sent by the individual library, with Elsevier filling in some holes (the sharing of use data within Consortia is permitted). Ideally we would have used data from at least 2 years, but SD was not COUNTER-compliant until 2003. I strongly recommend that at least 2 years of data be used by others who are seeking to build a shared title list.

Removing subscribed titles. Institutions are locked in (or 'guaranteed') ongoing electronic access to their subscribed title list. One way to limit the size of an STL is to remove these titles from each institution's data, allowing us to focus on leased title (Subject Collection) use. Subscribed title lists with ISSNs were readily available, and Microsoft Access 'Find

unmatched query wizard' was used to remove each institutions subscribed title data from its overall electronic use data.

Looking at the Subject Collection use data. Use distribution patterns formed 3 rough groups (Figure 1a-c), with each set conforming well to Zipf's Law: the majority of each institution's views were from a minority of their titles. Group (a) consisted of three schools that had much lower overall use, reflecting their smaller size. This showed that they would need a smaller, more affordable list. Elsevier had ignored their holdings when constructing the UTL proposed to the Consortium, and had made no accommodation for their smaller size.

Group (b) had more titles in the 10-100 view range, yet still noticeably fewer than group (c). This can be seen in the still negative exponential curve in (b) versus the more sigmoidal curve for (c). These two groups lose their distinction, however, when the number of titles used 0-10 times is considered. It is interesting to note that Pepperdine users seemed to use a wider variety of titles than their peers, as evidenced by a shift from clear membership in group (b) in their higher use titles, crossing over to group (c) type use for titles used less than 10 times. This anomaly may be explained by the impact of a series of Elsevier-led training sessions (held only at Pepperdine in 2003) that encouraged use of ScienceDirect as an A&I database. Finally it should be noted that on average, more than half of each institution's subject collection titles went completely unused in 2003 (Fig. 1a-c; mean = $50.8 \pm 4.9\%$ SE).

Table 1. Description of smaller schools in the SCEL Consortium ScienceDirect contract. Subject Collection Titles indicates the number of unsubscribed titles in the leased collections previously held by each institution.

Institution	Subject Collection Titles	Subject Collections
Mount St. Mary's College	575	4
Mills College	445	4
California Lutheran U.	370	3

Determining inclusion criteria. Graphically described, the task was to fairly determine where to truncate each distribution and which titles made the cuts. The smaller schools had leased Subject collection access to between 370 and 575 titles (Table 1). Separate spreadsheets for each institution allowed determination of: a) percent of total views in each title and b) cumulative percentage of views in highest use titles (after sorting from highest to lowest number of views, e.g. Table 2). Analysis of each of the three schools data showed that a 53 title shared list could be certified to include either: (1) every title used at least once a month at any of the three institutions, or (2) all of the most frequently viewed titles representing the top 66% of use at each institution, whichever included more titles. One institution requested the addition of 2 titles to this list, and the resulting list priced out at less than the cost of 2 subject collections. It was agreeable to downsize to this degree, because this group was particularly concerned about shrinking their Science Direct commitment (which for schools of this size consists mostly of their leased title holdings).

Table 2. A sorted use list with cutoff point (triple line). For this institution, cumulative use of 66% was satisfied by the top 6 titles, which included all titles viewed an average of at least once every 2 months.

TITLE	ISSN	2003 e-use	% of total views in title	Cumulative use
Animal Behaviour	'00033472	41	25.95%	
Epilepsy & Behavior	'15255050	28	17.72%	
Brain and Language	'0093934x	13	8.23%	51.9%
Contemporary Educational Psychology	'0361476x	10	6.33%	
Cell Biology International	'10656995	7	4.43%	
Biochemical and Biophysical Res...	'0006291x	6	3.80%	66.5%
Journal of Experimental Child Psych...	'00220965	3	1.90%	
Religion	'0048721x	3	1.90%	
Biological Journal of the Linnean Soc...	'00244066	2	1.27%	
Brain and Cognition	'02782626	2	1.27%	
.....	
Total		158	100.00%	100.0%

The larger schools' inclusion criteria were more complicated, even for the first cut. Subject collection title number ranged from 600 to 950, and equitable treatment required a compromise (Table 3). The first three schools contributed every title viewed more than once a month, as before. This covered 68% of their total subject collection views (on average). Applying this criterion to the last 4 schools would have over-represented their collections (i.e., covering an average of more than 91% of their views). To compensate, I doubled their minimum use criterion to include all titles viewed more than twice a month. Although this would appear to be a disadvantage to these schools (in terms of absolute use per title), it was still the case that a greater percentage of their total subject collection use was represented (Table 3, avg. 82% vs. 68%). Under this scenario, 766 titles were selected, yielding a deduplicated list of 400 titles (we'll call this list A). The base price of list A was equivalent to about 5.5 subject collections. So the quality of this list should be improved over subject collections, but it did not provide significant savings.

Table 3. Criteria for Shared Title List inclusion of higher use titles from the 7 largest schools in the SCCLC ScienceDirect agreement. Subject Coll. Titles is the number of unsubscribed titles in the leased collections previously held by each institution, which defines total use for the inclusion criteria values. Inclusion criteria columns refer to the set of titles viewed more than once or twice a month at each institution, with values indicating the percent of total use covered by that set, and (parentheses) indicating its number of titles. Shaded boxes indicate the criteria that were selected for each school. [Bracketed percentages] indicate the percent of total views that would have been included had the alternate criteria been used.

Institution	Subject Coll. Titles*	Inclusion Criteria	
		>12 views	>24 views
Loyola Marymount U. †	900	64% (45)	[40%]
Pepperdine U. †	800	66% (117)	[44%]
U. of San Diego	950	74% (101)	[57%]

Western U. of Health Sciences	600	[87%]	74% (78)
Claremont Colleges	850	[90%]	77% (146)
Loma Linda U.	700	[94%]	85% (133)
U. of the Pacific	750	[93%]	85% (146)
Average >>>	787		76% (109)

* Rounded to the nearest 50

† Two schools chose to keep some of their subject collections because they had already bought corresponding backfile collections. Data from these subject collections (for these two institutions) were removed prior to analysis and list building.

List A was based on individual institution needs and use patterns. It seemed appropriate to base the second cut on SCEL C-wide numbers in order to balance individual needs with the value of titles to the consortium as a whole.

Thus I determined the SCEL C-wide price per view for List A based on 2004 price (0.25x% list / 2003 combined views). Although all 400 titles met our criteria for 'high' use (at least once a month by at least one institution), more than 20% had a price-per-view greater than \$10 (see Table 4 for those over \$20). Their median list price was \$3570. Placing these 87 titles on the 'chopping block' allowed our consortium to play a small part in fighting journal inflation by cutting support for as many high-priced titles as possible. This cut (had it been permanent) would have brought the STL to 313 titles with a base price equivalent to 3 subject collections (called List B).

Table 4. The 31 highest price-per-view titles for the SCEL Consortium. Number of views is intentionally hidden to avoid revealing specific terms of the contract.

Title	2003 SCEL C views	2004 list price	Total SCEL C price per view
Archives of Biochemistry and Biophysics	#####	\$4,578	\$61.63
Journal of Materials Processing Technology	#####	\$4,701	\$58.76
International Journal of Food Microbiology	#####	\$3,282	\$41.03
Journal of Organometallic Chemistry	#####	\$10,545	\$36.91
International Journal of Biochemistry...	#####	\$2,873	\$35.91
Carbohydrate Polymers	#####	\$2,750	\$34.38
Chemical Physics Letters	#####	\$11,750	\$33.17
Marine Geology	#####	\$3,375	\$32.81
Chemical Geology	#####	\$3,627	\$31.74
Remote Sensing of Environment	#####	\$2,496	\$31.20
The Journal of Steroid Biochemistry and ...	#####	\$4,659	\$29.12
Clinica Chimica Acta	#####	\$5,097	\$27.87
Agricultural and Forest Meteorology	#####	\$2,497	\$27.31
Insect Biochemistry and Molecular Biology	#####	\$2,045	\$25.56
Developmental & Comparative Immunology	#####	\$1,872	\$25.20

Title	2003 SCELC views	2004 list price	Total SCELC price per view
Peptides	#####	\$3,598	\$25.19
Mutation Research...	#####	\$4,314	\$25.17
Polymer Degradation and Stability	#####	\$3,732	\$25.12
Comparative Biochemistry and Physiology...	#####	\$3,555	\$24.89
Vision Research	#####	\$3,250	\$24.73
Journal of Immunological Methods	#####	\$5,088	\$24.06
Palaeogeography, Palaeoclimatology...	#####	\$3,571	\$24.04
International Journal of Production Economics	#####	\$2,002	\$21.90
Journal of Aerosol Science	#####	\$2,247	\$21.85
Thermochimica Acta	#####	\$8,591	\$21.79
Journal of Chromatography A	#####	\$12,344	\$21.60
Journal of Banking & Finance	#####	\$2,331	\$21.47
Journal of Financial Economics	#####	\$1,881	\$20.57
Biochemical Pharmacology	#####	\$6,603	\$20.27
Plant Science	#####	\$3,795	\$20.13

The response to the List B proposal was mixed—some argued we should cut more high-priced titles; others contended that we needed to add the majority back. In the end we agreed on an intermediate value that each institution could add back, and it was up to each institution as to whether to request restoration of some of the 87 titles cut from List A or addition of other titles. The final list (C) contained 425 titles (more than List A!), but at a base price equivalent to 5 subject collections, demonstrating success in bringing down the average price of titles (vs. List A which had 400 titles for the price of 5.5 subject collections).

Evaluation of the Shared Title Lists

STL adoption and benefits and costs relative to Subject Collections. Seven of nine institutions chose to replace their subject collections with a shared title list. All three of the schools with lower use overall elected to lease the 55 title list. This STL allowed them to cut their lease payments by at least half in return for relinquishing access to 9 out of 10 of their subject collection titles. This may seem extreme until it is realized that 60-70% of their titles were not used even once in 2003.

Five of the seven larger schools selected the large STL. Although it represented a 30-80% decrease in the size of an institution's leased title collection, its potential impact on each large institution's bottom line was much more variable than for small schools (Table 5). For two of the seven, it represented a sizable *increase* in price over their subject collections. Perhaps not surprisingly, one of these two schools chose not to lease the STL, citing the recent addition of 2 more subject collections, increasing use in 2004, and the feeling that subject collections were sufficient to cover their relatively narrow range of needs. The other school that decided against leasing the STL tied for the lowest price decrease, had bought backfiles of at least one subject collection, and perhaps most importantly, could not be convinced that less is more (i.e. that fewer higher quality titles is better than many more with very low use). Zipf's Law did (of course) hold for large institution subject collection use as well: 30% to 70% of each large institution's leased titles were not used in 2003.

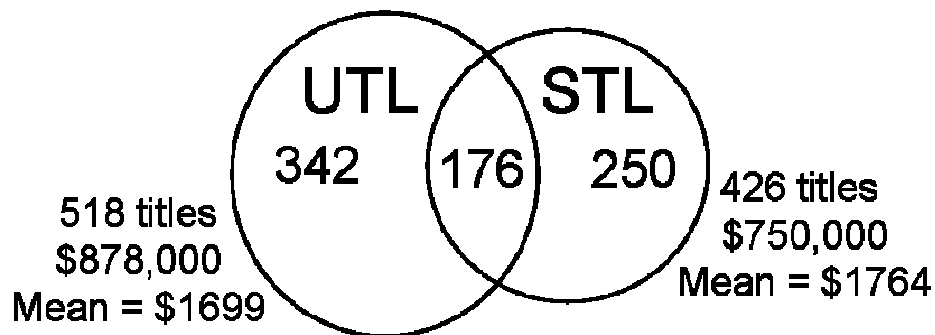
Table 5. Shared title list benefits and costs. Shaded rows indicate schools that selected the STL. Overlap with subscribed titles indicates the number of titles that appeared both on the STL and the school's subscription list. Credit for overlap is the percent of STL base price saved, and savings versus subject collections compares each institution's STL price after overlap credit to the price they would have paid for their subject collections.

Institution	Overlap with subscribed titles	Credit for overlap	Savings vs. subject collections	% fewer titles vs. subject Collections
Small List	(Total of 55 titles)	--	--	--
A	1	0.2%	61%	91%
B	0	0%	61%	88%
C	1	0.2%	48%	85%
Large List (C)	(Total of 425 titles)	--	--	--
D	175*	48%	59%	70%
E	275*	47%	45%	79%
F	75*	18%	35%	49%
G	25*	5%	13%	58%
H	25*	5%	13%	55%
I	50*	20%	-26%	47%
J	25*	17%	-31%	29%

* Rounded to the nearest 25

Difference from UTL. Another subjective evaluation of the STL is to ask whether it is notably different from the SCEL C UTL offered by Elsevier. Our expectation that the set of shared subscriptions in the UTL would *not* contain the titles that were most used at other institutions was borne out: the two lists had fewer titles in common than were unique (Fig. 2).

Figure 2. Venn diagram comparing the subscription-based Unique Title List offered to SCEL C by Elsevier to the use-based Shared Title List that was designed as an alternative. Dollar values indicate the total and mean 2005 list price.



Choosing quality over quantity. The STL has nearly 100 fewer titles than the UTL we were offered, with a similar average list price. On average, adopting the STL meant loss of access to more than half of an institution's leased titles (58%). Although those who see this as a significant loss can be easily countered by the observation that an average of 50% of subject collection titles had zero use, this still lessens the extent to which users and accreditation agencies can be impressed by the large total number of subscriptions an institution has access to. Indeed, in every case the percent of titles that became inaccessible (after replacing the

subject collections with the STL) exceeded the percent cost savings. The challenge I believe libraries must rise to is that of interpreting this as an appropriate choice, because it enhances our collection quality and may put pressure on Elsevier to maintain the quality of its journals (rather than simply constantly increasing the quantity).

Future assessment. Objective tests of the effectiveness of the shared title lists can be performed after a year of use of the new collections. Do the new lists decrease the number of titles getting no use? Is there a lower individual or SCCLC-wide cost per use for leased (STL) titles? In addition to these questions, it remains to be seen how well these lists will adapt to the changing needs and new SCCLC ScienceDirect members, but they are almost certainly more flexible than the UTL, and subject collections are still an option. Additionally, it might be interesting to examine impact factor per dollar of the STL versus a set of subject collections.

A clear benefit. One thing is certain: these lists have restored the ability of the SCCLC libraries to add individual Elsevier titles in the future. No longer will we need to make excuses (though they may have been valid) about not being able to add access to high-priced ScienceDirect titles because they can only be purchased in packages or at full list price. Shaving off the lowest use titles and replacing them with more desirable ones is second only to an unlimited budget, and may even bring a touch of efficiency back into this market!

Conclusion

It's not the number of titles that counts, but their quality and value to local users. Sharing use data to build the optimal shared access list represents one way consortium members can work together to improve the quality of a product for their users. Library vendors (even some of the most notorious) are willing to work with consortia to increase the value of their products—even if it means at least a temporary loss to their bottom line. Collection development, serials, acquisitions, and electronic resources librarians must take issue at the least attractive pricing models and negotiate with publishers to develop viable alternatives in order to be effective in a world where things are seldom what they seem.

Reference

Ives, G. 2005. Content overlap of the Elsevier ScienceDirect Subject Collections: book series, current journals, and journal backfiles. Contributed paper from the 9th World Congress of Health Information and Libraries. Retrieved on 27 Jan 2006 from: <http://www.icml9.org/program/track2/activity.php?lang=en&id=7>