

Presentation Script - Melissa Belvadi - APLA 2013 - Cultivating a Sustainable Collections Garden

[Note: this is my script as written, but in the live presentation I ended up deviating from it quite a bit for rhetorical and timing purposes.]

[title slide]

Hello,

A couple of weeks ago, our ILL technician came to me. She had received a request for a book that we would have to go out of our regional ILL consortium to get, so it was going to take a couple of weeks to arrive. She had also gotten from the same patron an email that strongly recommended that we actually buy this particular book.

As it fell into my liaison area, I looked into it and agreed, but more than that, found that in terms of filling the patron's ILL request, we had not one, but three different options, as the book was available as an ebook, which I could obtain within about an hour for him, or from Amazon.ca, which we could obtain as fast as 2 days depending on how much we wanted to pay for rush shipping. Three choices, depending on the tradeoff of service speed and price.

That is, the concept of "just in time" document delivery for this item was no longer confined to ILL, but now involved collections options.

Historically, academic libraries have been primarily in the "just in case", not "just in time" business. Many libraries, like UPEI's, had their basic collection development policies and budget allocations set long before the existence of ebooks or Amazon, indeed long before online anything. ILL and book purchasing were, by today's standards, very slow.

Many libraries didn't even offer ILL service to undergraduate students, at least not for free, so, unless they lived in driving distance of other universities, they would be pretty much restricted in their research to what their own library already had in its collection.

Further, the library had no easy, reliable way to collect data on what books and journals were actually being used, as circulation was a process involving stamping numbers on cards, and journals didn't circulate at all, so only in-house re-shelving stats were available.

And of course, like today, almost no library could afford to purchase or subscribe to anywhere close to all of the scholarly output available across all of the subjects taught and researched at their institution.

In such a world, a couple of practices made a lot of sense.

First, librarians needed to carefully select what titles to buy. Typically, librarians would be assigned one or more subject areas to develop particular expertise in, for both collection development but also for research instruction and assistance.

Second, budgets, which are numbers, had to be allocated somehow, and libraries started to use proxy data to do so. You're probably used to the word "proxy" to mean a server for providing

off-campus access, but I'm using it in the more traditional way, in the way APLA uses it this week to handle absentee voting - using one source of data to stand in for the actual data you can't get, hoping that the data you can get correlates well with what you can't.

So when it comes to allocating library collection funds, what data was available in the pre-online world that might have a good correlation to what we actually want to know about patron demand for books and serials?

[slide - Allocation Formulas]

Common proxies include number of student majors in each department, or even total credit hours taught in each department, number of faculty in each department, etc. Libraries serving a mix of undergrad and grad students might numerically "weight" graduate enrolment or credit hours more heavily than undergrad.

They would develop complex formulas, and bring into those formulas additional weighting factors like how much average books and serials cost in a subject compared with others, and what can only be "fudge" factors like a weight to take into account that some disciplines are more "bookish" as one author of a formula-method article called it, than others.

I call this a "fudge" factor because there's no objective basis to setting such weights for the particular institution, so such factors can be used to force the results to a pre-decided outcome.

Using proxy data like this might have been necessary in the past, but we know that it caused some really big problems that many of us are familiar with even today..

[slide - Allocation Divisions]

First, how many of you have your book budget divided by departments that closely map onto your institution's academic departments, as opposed to some other kind of subject categories like LC class?

How many of you do that for serials/subscriptions as well as books?

Those of you who do, would you say that the faculty in those departments take a keen interest in how much "their" budget is, and want to know how it compares to other departments'? That is, does this allocation method result in advocacy politics?

Now the painful question, which I'll ask rhetorically - do you find that the liaison or subject librarians tend to internally reflect that advocacy politics - is there a sense of competition among the librarians themselves for greater shares of the pie, especially when budgets are cut and the pie shrinks?

[slide - UPEI's Allocation System]

The librarians at UPEI, long before I started there, switched from a departmental-centred

allocation to one based on LC Classes. Some of those classes do map obviously to some departments, but we discourage such thinking, even to telling external program review and accreditation groups that we simply can't provide them with such numbers when they ask for them, and explain why. You may be interested to hear that we've never been criticized in their reports for that failure.

We also use the LC Class allocation method just for books, not subscriptions. That decision also predates my being at UPEI, but I believe it's because it has become extremely difficult with the Big Deal packages and the huge multidisciplinary aggregate databases like Academic Search Complete, to meaningfully divide the subscriptions budget by subject.

However, because the liaison librarians do deal with departments, not LC classes, when it comes to library instruction, and we are encouraged to build ongoing working relationships with "our" faculty, we do still have an advocacy, or what I call the "going native", problem when we discuss our collections budget.

[slide: Evidence-Based Budgeting]

Unlike in the past, however, we now have for the overwhelming majority of our expenditures, high quality data on what patrons actually need and actually use. I'm talking about usage and circulation data in the case of what we already own, and ILL and other sources of data, for evaluating potential new purchases and subscriptions. The number of cases in which proxy data is needed is smaller and smaller every year and the kind of proxy data available is also much better than the old institutional head-counts.

This then, is "evidence-based" collection development budgeting. Decisions based on hard, verifiable, usage data are much more impervious to departmental politics, and much more likely, I believe, to be accepted by faculty, many of whom use quantitative data themselves in their own research.

I'd like to show you now several examples of what we're doing at UPEI to apply usage data to collections budgets.

This work arose first out of a budget cut we had last year, when the collections budget had to absorb an approximately 8% nominal cut, plus inflation on top of that pushing the effect well over 10%. At the same time, we realized that we had no useful mechanism for adding new subscription titles that our data showed high demand for. Then we were faced with the possibility of an even larger cut this year, as HST came to PEI, and an at best flat nominal budget meant inflation was going to do what it does every year as well - the two would result in an approximate 13-14% cut on top of the previous year's cut.

After years of 0% budget increases, the librarians had unofficially accepted that annual subscription price increases were eating into the book budget, but it wasn't really recognized in our budget process very much. The librarians tended not to spend more than 50% of "their" book budgets before Christmas anyway, and by then we had the biggest subscription bills

behind us and had a good idea how much money was left, and just did an across-the-board resizing of the book subfunds to match the remaining money.

So we don't have two budgets - books and serials, but just one, and we needed a way to evaluate effective and ineffective spending across the one budget and make some hard choices.

[slide - Budget Tradeoff]

If we're going to analyze our budget as one big budget, we have to get past the "apples and oranges" problem of what "usage" data means for the print book collection. By the way, when I say "print book collection" the same subfunds cover individually purchased ebooks but since they're still the vast minority, it's easier to talk about it as the print books.

[slide - Print Books as Subscription]

So I introduced the idea of the print book collection as being analogous to a serial subscription where you accrue perpetual access rights each year to each volume you buy. Viewed this way, using circulation transaction data, we can actually calculate a "cost per full text use" for the physical book collection.

Here's what it looks like:

[slide - UPEI Circulation Data]

I pulled the circulation transaction data directly from our Evergreen system, and put it into Excel to do this and some later calculations.

It's fairly simple conceptually - take the amount you spent on books in those 5 years, and divide it by the number of times all of your books were checked out in those five years. It's quite comparable to online journal stats where you are given the total number of uses of that journal, but not broken down by which year of the journal, so the data includes all of your "backfile", not just use of the current year's volume of the journal.

I pulled the circ data for the entire circulating book collection (except for the very old stuff in compact shelving that has such low use that it just throws the ratios off), and then again for just those books with a 2008+ pub date. My available circulation data runs from June 2008 to present. The line here about "no circs" isn't really part of this calculation but it was one of the facts I could get during the same analysis, and I'll refer to it later.

So now we have a hard number, \$8.42, or let's just say \$8 so as not to imply a false precision, per use of the annual "subscription" that is our book budget.

[slide - Cancellation analysis]

So we then went through a process of examining every electronic subscription that costs \$500 or more and compiling its cost per use.

We did only the electronic ones this year, because we did a thorough review of the print ones the previous year, and they are the ones that don't have clear hard data but still require proxy data, which I can talk about if you interested at the end, and that's very long and tedious per title, and they're really pretty much under 3% of our budget now, a large chunk of which is a handful of "must have titles" - Science, Nature, JAMA.

I limited it to \$500 and over just because of time limitations, because we were going to go through a line by line process with all 7 librarians present so there would be no question of acceptance or concern about discipline overlap and someone getting left out of an important cancellation decision.

So here's a piece of what that spreadsheet looks like:

[slide of possible cut analysis - price column redacted for website version]

I'm sorry this is hard to read. This is a section of the list. The columns are the title of the product, a generated link to our holdings for that title which is useful if it's a single journal, the price, the current subscription period we've already paid or committed by contract through, and then the usage analysis. "cpf" means cost per full text use.

So for instance, Nursing Clinics of North America has a cpf of about \$13. Project Muse has a cpf of just under \$5. PsycArticles has a cpf of \$1.39, but Physical Review Letters has a cpf over \$50.

Remember, we have a lot of money to cut from our budget, and there's only one budget, not two. The reasoning is that anything that had a worse cost-per-full-text-use ratio than about \$8 should be cancelled, and better than that should be kept, which given a fixed cut target means, that equivalent amount might be cut from the book budget instead.

[slide - Usage Data Sources/Issues]

The overwhelming majority of products comply with the COUNTER standard.

Is there anyone here who isn't familiar with the COUNTER standards?

[If so, explain briefly - standardized definitions and presentation rules.]

I have an assistant who pulls the calendar-year COUNTER reports product by product. I'm looking into how to use the SUSHI functions of our ERM, but haven't gotten there yet. Is anyone here using a SUSHI system or any commercial statistics-gathering service to retrieve your usage data for you? [if so - I envy you! - we'll get there eventually]

For those not familiar with it, SUSHI is a data standard, a kind of API, that vendors can choose to support to provide their COUNTER-compliant data reports directly into customer software, rather

than what we do, which is manage a long list of vendor admin accounts to which we login and manually download our COUNTER reports.

Of course, when you embark on a list like this, you start to find all kinds of oddities and special cases. The biggest one was A&I indexes - there's no "cost per full text use" so we looked at cost per "abstract viewed" instead, with an understanding that that would have to have a much lower threshold than \$8 to be justified. As it happened all of our A&I indexes were so low in their cost per abstract viewed that it was never an issue. For instance, PsycINFO is just 20 cents per abstract viewed.

Several products either don't make sense at all to think in terms of full text, like Naxos streaming audio, or would not report anything comparable to "full text documents viewed" but would force us to accept only "number of sessions" or "number of searches". Again, for all of these products, we found the usage data they did provide was so high relative to cost, we didn't need to come up with bright-line thresholds, but just agreed to keep them.

[slide - Non-usage factors]

If you're being evidence-based, you should avoid any extraneous, subjective criteria, or even objective criteria that are irrelevant to your institution. An example of a relevant non-usage criterion is whether the resource is absolutely required for program accreditation. We did a survey of our programs and found that not a single one could identify a specific title of a journal or database that was absolutely required, although chemistry comes close with quite a lot of wink-wink nudge-nudge regarding certain ACS titles.

We also considered very rare exceptions for extremely expensive journals that had such high absolute use rates that it would be a burden on ILL to absorb, as well as questionable in terms of copyright law whether we would truly be "substituting" ILL for a subscription, even though the relative use rate, the cost per use, crossed our cancellation threshold. An example is Proceedings of the Royal Society B, which has a cpf of \$14, but has about 250 uses per year.

What should NOT be a criterion is asking the faculty if a title is "important". If a title is important, that importance should be showing up in the usage data. Remember what I said earlier about librarians "going native"? In our first round of this kind of analysis, one of our librarians kept trying to protest that this or that journal title which had usage below 10-20 per year was "important" or a "core title" in the field and should be kept on that basis.

For this process to have integrity, though, and everyone to perceive it as fair, it's important that everyone agree to the thresholds in advance of seeing the title-by-title data, and important to have all of the librarians, or if you have many more than we do, at least many of them representing different disciplines, in the room at the same time, so it doesn't become a one-on-one argument between the collections librarian and a single liaison librarian.

That was a mistake I made in my first year doing this, thinking I wouldn't waste librarians' time having them present to discuss titles outside of their subject areas. It turns out to be very necessary to keeping the process "fair".

There is one criterion that seems at first to have quantitative authority unrelated to usage, and that's ISI impact factor. However, unless you are collecting for a special reason beyond day-to-day use by patrons, as we do at UPEI for our PEI Special Collection and our Veterinary Collection, the international impact factor isn't as relevant to your budget as it at first seems.

Remember that for journals even more than books, lack of immediate access is not a serious barrier because ILL is so fast today. If it's not being used much, don't subscribe.

[slide - Proxies for Usage Data]

A couple of products simply refuse to provide any usage data at all such as Abstracts in Anthropology and the National Council of Teachers of English journals package online. I contacted both vendors, and was told "sorry, we can't give you any usage data at all".

Ok, so what do you do for those vendors who don't give you data? It's not fair to just keep them for lack of data. So we have to look for some kind of proxy data for these.

Here are some proxy data sources that you can use that I think are probably much closer to getting an accurate view of reality than the traditional institutional-counts I talked about earlier.

[slide - Ezproxy server data]

One of the very best is your ezproxy server logs, if you have ezproxy. Every use of the link to this database from the library web site, on-campus or off, puts a line in our proxy server log files. We have the log files and can count these as "sessions".

[slide - Cancelling an index]

For example here's the case of Abstracts in Anthropology, which costs us just under \$1,000 per year. Because it's an A&I index, not a full text source, and my data is proxy data, not direct usage, I wanted to have as much relevant information as I could find for our faculty - first, is it being used much, and second, do we have other ways of searching the same content it indexes?

Using the proxy server logs, I determined that there were only 82 uses, which comes to almost \$12 per session. Admittedly, if the faculty have created their own direct bookmarks to the database on campus, bypassing our proxy server, we won't count those. This is partly why part 2 is also important.

I had a student assistant help me with part 2 - indexing coverage overlap analysis.

First we identified the most important journals in the field. We did this using a combination of resources, including, yes, impact factors, as well as articles from the library professional

literature about research in this field and other academic libraries' lists of core titles. We tried to err on the side of being overly expansive, even though we know that our anthro department is not much focused on physical anthro, just social anthro. But for that reason, we didn't feel the need to include in our analysis the thousand purely medical journals that A in A indexes highly selectively, which are all indexed in Medline anyway.

[slide - Abstracts in Anthro coverage]

Sorry again for the hard to read slide. The columns are the journal title, whether it is Core, the link to our holdings, what platforms if any we have the title in full text, whether it is indexed in Academic Search Complete, whether it is indexed in SocIndex, A search link to Google Scholar followed by the count on number of hits in Google Scholar, a search link into our EBSCO EDS system followed by the number of hits that produces and finally the range of years covered by the EDS results.

Our result here shows that ALL of the core titles are well indexed in EDS, and almost all in Google Scholar too. All but two (and those two are among the "least core") are also full-text searchable in one of a small handful of full text providers like Wiley and JSTOR. Most are also indexed in ASC and SocIndex, but not enough for either of those databases to be recommendable to be the sole substitute for Abstracts in Anthro by itself.

Based on the combination of the evidence in part 1 and part 2, I am, as the anthropology liaison librarian, entirely comfortable with going to the anthro faculty to justify this cancellation and advise them as to what to use in replacement and advise their students to use.

Ok, so that's subscriptions, at least when it comes to considering cancellations. We'll come back to subscriptions at the end when we talk a bit about additions.

[slide - Book subfund allocation]

Back to the book budget. Now that we've cancelled what we can agree to on the subscription side, we can look at what's left of our budget that can be applied to purchasing "one time purchases", and figure out how to allocate that among our LC class ranges.

Remember that data I showed you about circulation usage across our collection? Well, when I pulled that data, I actually pulled every single transaction line by line from our system since 2008, and included in each line its LC call number.

With not too much effort considering the amount of data involved, you can end up with a table like this:

[slide of LC Class breakdown]

This is the same circulation data I used in my earlier slide to calculate the \$8.42 per use, only broken down by our LC-based subfund groupings.

Remember that I ran the data twice, once for the entire book collection, Stacks plus Reserves that have an LC (so not including faculty copies), and then again but only including books that have a pub date of 2008+ so we can see if there are more recent trends different from the historical collection, the “backfile” if you will, although excluding the pre-1970 extreme backfile that we store in compact shelving.

What we need is a way to divide 100% of a budget across these LC class groupings, and this gives us just what we need. Each column gives percentages that can be used against the amount of money we have left for the “one time purchase” budget. In the original spreadsheet, we have decimals for a bit greater precision, especially those 0%s.

After that, I’m inclined to use the 2008+ data rather than the overall, but I could see an argument either way. But I would encourage the librarians to make the decision before seeing this table, so they make it on principle and not based on liaison area advocacy. Both columns add up to 100%, very easy to convert them into number of dollars to be spent.

Of course, the librarians are free to agree to set aside particular amounts of money for specific reasons before determining what the amount is that will constitute the 100%. For instance, we might want to pull \$50,000 from the pot to fund a DDA program separately from this spending - more about DDA later.

[slide - Subscription Additions]

So, enough of depressing subscription cancellations and book budget cuts.

What about evidence-based subscription additions?

Two years ago, after a round of cancellations of very low-used journals, we thought it was time we look at the available data to see if we were neglecting any journals that we should have, but don’t. There were three major sources of evidence for neglected subscription titles to look at: ILL and the log files of our openurl resolver, Godot. ILL is obvious, but Godot is the system that records the citation they were looking for when clicking on our “Check@UPEI” button from our citation databases when the full text is not already there. And both EBSCOHost and Proquest, our most important A&I platforms, give us title-level lists of number of abstracts viewed.

These are in increasing order of strength of user interest, as patrons look at many abstracts casually, only bother to click on the Check@UPEI button for some of them when the full text wasn’t immediate present, then if Godot tells them we don’t have the title, even fewer will bother to complete the ILL form, which is part of Godot.

We looked at the last three years of data in each, and found a handful of titles which were posting pretty high stats in both the Godot and ILL lists. Unfortunately, of the top 6 which stood out from the rest of the list, 4 were all nursing-related, and the other two on the list which weren’t were utterly unaffordable.

Fortunately, by that point in the process, the librarians had accepted the evidence-based

methodology well enough not to be too bothered by what in a more advocacy-based process would surely have been a sense of “unfairness”. So we went ahead and added the subscriptions for the 4 nursing journals.

I expect that we’ll do a similar analysis every couple of years to see what’s bubbling up to the top of the data, and make sure we aren’t missing anything important that we can afford.

[slide - Book Additions - Evidence-Based]

Now, can we use evidence-based methods for book additions as well as subscription additions?
Yes.

First of course, there’s patron driven acquisition, also known as demand driven acquisition. We have a DDA program in place with YBP for a few specific subject areas, but most of the librarians have been hesitant to jump into it yet. We actually did make an attempt to do a broad DDA with EBSCO in October, but after we spent \$3,000 in 5 days, of what was supposed to be a \$6,000 pilot, we had to hastily cancel that plan, and haven’t gotten back to trying anything that ambitious again yet. That was just a few weeks after we had gone live with EBSCO’s EDS discovery product, though, and it made us very aware just how effective EDS is compared with our catalogue at exposing ebooks to our users. I did my presentation at last year’s APLA about DDA, so I’ll leave it there for now.

Then of course, there’s ILL book borrowing data. Remember my opening anecdote about the three options to obtain the book the patron wanted?

[slide about ILL distance program]

In two weeks, we have our annual Collections Budget meeting, and I’ll be proposing a new procedure to allow ILL to work directly with our acquisitions technician, to begin to routinely make such purchases without seeking liaison librarian approval per title, subject to well-defined price and availability constraints. By its very design, in addition to enhancing our own collection, this program would be purchasing books not available in the rest of our local ILL consortium, a kind of unilateral collaborative collection development on a small scale.

[slide- Book selection - who do we trust?]

There’s an interesting disconnect between serials and books among librarians. Librarians who have no problem at all with determining which serials to subscribe to on the basis of demand use data suddenly have a big philosophical problem with doing the same for books. One worded it to me this way: “I’m not sure that buying a book just because one student wants it now is the best way to develop a collection.”

I thought about that a lot, and here’s my philosophical response first. Who do you think is more likely to predict correctly what books next year’s undergraduate student in major X is going to need? This year’s undergraduate student in major X, taking the same classes that the future

student will take, doing likely the same assignments under the same syllabi or very similar ones? Or a middle-aged librarian who probably didn't even major in that field, or did so 30 years ago at a different institution, and barely has time, at best, to skim through the latest syllabi for the courses being offered now?

Now my evidence-based response. If we buy a book because a specific student requests it now, we know for certain that the lifetime use count on that item will be at least 1.

In contrast, as my much earlier slide showed, 46% of the 2008+ published books the librarians have hand-picked (and we don't have an approval plan, so these were all individually hand-picked) have no circulation transactions at all so far. Other studies at other institutions that have been published show similar results, and further studies also show that if a book doesn't circulate at all in its first 5 years, it very likely never will.

So I'll literally put my money on what the patrons tell me what they want.

[slide- Thank you.]

Thank you.

Any questions?

I