The Swiss Cheese Problem

Jacob Nadal, Executive Director ReCAP
http://recap.princeton.edu

Presentation for ALA Annual 2014, Book and Paper Interest Group Special
Session: "Space Oddity"
Presenters were asked to prepare short, humorous talks on serious problems in
storing library materials.

[Start]
The Swiss Cheese Problem, or, Chesterton Redeemed: Being a response to
and enlargement of "The Neglect of Cheese in European Literature" through a
comparison of the formal qualities of selected foodstuffs to problems in
Library Economy.

[Slide 2]
A cheese. Observe its full, dense, and well-compacted flesh, the even
disposition of the central, nourishing cheese-stuff, surrounded by a firm and
durable rind, useful in the aging of said cheese and preservation of its
wholesome qualities over time. The analogy between this ancient foodstuff
and our own plan for the high density storage of library collections should be
obvious: a dense and consistent arrangement of the vital content with a
protective wall isolating said content into suitable environment for the passage
of time.

[Slide 3]
Now, colleagues, I present to you a cheese of the variety known in the
American vernacular as "the Swiss", and although connoisseurs will balk at
this vulgar amalgamation of pedestrian cheeses with such noble comestibles
as the Emmentaler and the Jarlsberg, for the purposes of Science and of
Economy, I tell you these cheeses share alike a tragic flaw, exhibiting a
profusion of non-fromagenous zones filled with a wastrel ether, such that they
are volumetrically inferior to other varietals whose more uniform packing
makes them, cubic inch for cubic inch, the cheesier.

[Slide 4]
Were this a simple problem of household economy or dining under the constraint of space, such as afflicts the astronaut or the submariner, we could fondue and be done. Yet a like problem pertains unto our book collections, and while we may aspire within our repositories to match the consistency of the produce of the Helvetic Confederation - their trains, timepieces, and yes, their slurried national cuisine - the ponderous reality of our bibliographic storehouses precludes an easy solution.

[Slide 5]
For within these stacks, volumes are shelved like with like, sorted by their respective sizes into trays, which are thence divided into batches so as to fill each shelf plate to great planar efficiency, and then the shelves are stacked each above each to gain the utmost in vertical efficiency. The withdrawal of a single volume therefore leaves behind a gap of quite specific size, as difficult to refill as an eye in a Swiss cheese.

[Slide 6]
In the classic formulation of this problem, each gap must be filled in by a volume whose dimensions are almost exactly the same as the volume removed. A laborious and pedantic undertaking to be sure, more expensive in some reckoning than the cost of simply building additional shelves. For a facility the size of ReCAP, this process could pass over from the merely tedious to the nearly Sisyphean.

[Slide 7]
An analysis by our estimable colleagues at OCLC research suggested that within our collection, now just over 11.3 million items, approximately 10% of the materials were duplicate or triplicate copies. At this scale, the rewards of de-duplication would be great: to wit, space for 1 million additional items. Yet the cost and time required to perform this feet suggest that it is an utter folly, and that we must resign ourselves to heave the sigh, regret the errors of our early accessions, and move forward with more wisdom of purpose than we began.

[Slide 8]
But friends, colleagues, I ask you: Is this resignation meet and right for us? Shall we neglect the sacred calling of The Librarian, our duty to the increase

of order and perfection of array? Let us not mistake quietness for quietude! For the previous method of bibliographic infill was deficient in seeing only the book yet ignoring its surroundings and the many systems that placed it in its place. Colleagues, we shall look beyond the book to behold the whole of the repository! We fly towards an horizon beyond which lies the possibility of a more effective means to re-sort high density storage and optimize the placement of the items there within!

[Slide 9]
Now, behold two trays, containing a duplicative item. Since we desire to retain but one copy, we perform whatever comparison is needful to select the volume desires, and then one tray, (a), is returned to the stacks in precisely the same state as it left them, fully intact and unaltered. The remaining tray (b) is withdrawn from the inventory as a whole, the entire tray I say, removing the duplicate along with all of the other contents, and a new tray (c) is introduced. We now have removed the duplicate (b) from the repository, along with every other item in its tray, and added where we intend to subtract! Folly, you ask? I say thee nay!

[Slide 10]
For the contents of tray (b) are now divided into two trays -- the original (b) and a new tray (c) -- and then these trays added to the head of the accession queue. From here, the two partial trays are filled in with volumes newly arrived at the facility, and henceforth processed just as though they were new arrivals: accessioned, verified, and assigned to the shelves. With this system, we escape the prison of item-by-item thought! We free ourselves from outmoded and ponderous usage, breathe in the fresh air of the 21st century, and proceed to our long-sought conclusion in clearer, more modern prose. (Phew.)

[Slide 11]
The reason this approach is promising is that it makes use of existing workflows. So, rather than create a separate de-duplication process that creates gaps then requires us to infill those precise spaces, we make a small alteration to our retrieval process, pulling a tray rather than a book; then we do sizing as normal except into partly-full trays; and then just run the accession process as usual. This looks very good on paper - a cost of about $1 per item - because Accession is very fast on an item-by-item level. This may be hiding a big roadblock however. Reducing this process to item-based costs mathematically

we make an implicit assumption that we get free economies of scale in both directions (from item to batch, and from batch to item).

[Slide 12]
Sizing is the slow process in high density storage. No surprise, since that's where we actually go from disorder to order, taking many random sized books to groups of like-sized books. If we do the math at the slowest-batch level, pulling enough duplicates to create an hour's worth of sizing, and then derive an item-level cost from the whole batch cost, we get a much higher number, about $3. Why? Is math broken? No. You can see that if we eliminate the sizing costs, (see slides) we get right back to the $1 per book level we projected. That's how we isolated the mistake in the first place.

[Slide 13]
So, is all hope lost? No, and here's why: the sizing step may not actually matter in de-duplicating high-density storage. For one thing, we're starting with some pre-sized items from the trays. But at an even larger scale, analyzing ReCAP as a whole facility, we can look for high-yield trays with multiple duplicates and there may be hot spots with a large number of same-sized duplicate items. This sort of mapping can reduce our travel time for retrieving the trays and nearly eliminate the work of sizing materials. If so, when we look at costs of around $1.25 to $3.15 the "good-possible" and "worst-likely" scenarios, the project outlook is good.

[Slide 14]
For 442,422 monographs, the slowest type of item to de-duplicate, new shelving would cost about $2 million. That means that in our $3.15/worst-likely scenario, we can be over budget by 40% and still break even. Breaking even is worthwhile in capital planning, since it delays our next major construction project, giving us more time to accrue those funds. If things work well, and de-duplication is in the $1-2 per item range, we can save hundreds of thousands of dollars.

[Slide 15]
So what conclusions can we draw? In the near-term, there is a real possibility of de-duplicating high density storage in a cost-effective manner. On a more philosophical note, let us take seriously that we are library scientists, and not mere toilers in the stacks. Compared to the wages paid the average street urchin, student employee, or library clerk, the Librarian may seem a costly

thing. Yet the expensive hours they spend in consultation of their spreadsheets may preclude a rapid and unwarranted deluge of costs from a project ill-conceived, poorly tested, and ultimately ineffective. It is a little, well, cheesy, to wax poetical about one's profession this way, I know. But seeing as I'm talking to my colleagues at the American Library Association conference, I hope it won't grate too much.