

No-Growth Libraries and Citation
Analysis; or, Pulling Weeds with
ISI's Journal Citation Reports^{TM1}

June 30, 1975

Number 26

Weeding is an old-fashioned term in library management, but there is still no process so painful to librarians. I've known several librarians who could fire people more easily than they could discard books or journals.

Weeding a library is like examining an investment portfolio. Investment advisors know that people don't like to liquidate bad investments. How painful it is to realize that the hard-earned money you invested has not worked for you! Investment involves risk — and so does book and journal selection. Both types of risk-taking involve some speculation, some guesswork, and a bit of hard information.

Just like frustrated tycoons, many librarians can't face the fact that some of their guesses go wrong. They continue to throw good money after bad, hoping like so many optimistic stockbrokers that their bad decisions will somehow be undone. After paying for a journal for ten years, they rationalize, *maybe* someone will finally use it in the eleventh or twelfth year. It is difficult to buck the momentum inevitably gathered by a long-term subscription; it's so much easier simply to continue doing what has been done before.

In fact, the only object with enough inertial mass to stop this kind of irresistible force is an immovable budget. Weeding becomes essential during periods of stringency. Even in affluent times there is a limit to how much a library can buy. Nowadays the combined forces of information overload, inflation, and recession demand rigorous selection criteria.

This brings us to the concept of zero-growth for journal collections, first presented in a stimulating paper by Daniel Gore.² Libraries simply cannot keep on growing exponentially. The tacit assumption that everything published must be stored *somewhere* is no longer tenable. The Alexandrian type of librarian who thinks there is a mandate to collect everything must soon face the reality that a truly *complete* collection is not — and never really *was* — possible.

To aspire to collect everything is characteristic of an archivist, whose job is retaining materials that are seldom — if ever — used. Of course, the function of a library is different from that of an archive. Most libraries, as distinct from archives, have little if any demand for aged material — and when they do, they usually know what it is and have it conveniently located. Accessibility of material helps determine its frequency of use — and frequency of use *should* determine accessibility.

What has this to do with the nuts-and-bolts business of selecting books and weeding journals? Plenty.

The information explosion has come up against the inherent limitations of libraries — space, budget, and manpower — and the resulting collision demands that libraries no longer be thought of as archival storehouses, but instead be regarded as bibliographic search centers. This implies two basic changes in priorities for library storage: first from seldom-used documents to heavily used ones, and second from primary documents to all kinds of indexes and bibliographic tools.

Most libraries can afford to stress better bibliographic tools — even those that may be regarded as "too expensive" by traditional standards. For example, the cost of the *Science Citation Index*[®] (*SCI*[®]) may displace subscriptions to only a few dozen marginally used journals. But according to Williams and Pings,³ the *SCI* was in fact the "best investment" for hospital libraries.

Another change in libraries involves the process of weeding, or separating the seldom-used material from the heavily-used material. Let's first consider books.

As a book is cited each year, the probability increases that it will be used and cited again. If it has been cited 50 times, it has a high probability of being cited another 50 times. But, contrary to the intuitions of many librarians, if a book has *never* been cited it has a very low chance of *ever* being cited.

Keeping this in mind, ask yourself what kinds of books are likely to be donated to academic libraries. Are they the books departmental scholars frequently consult? Or are they the books the donors never use, but nevertheless find difficult to throw away?

The bitter truth is that the supposed benefactors of libraries probably *retain* the books that they — and others — want to use again and again. Donated books which are dumped onto libraries often merely clutter shelves and increase the cataloging load. The library would actually be better off with a second copy of a frequently-used book than with another book that no one will ever use.

Now let's consider journals.

According to the Bradford distribution, in almost any scientific field a small number of the *journals* publish a large percentage of the *articles*.⁴ Consider that 152 journals accounted for half of all reference citations to all scientific journals in ISI's 1969 study,⁵ and more recent data show that the same relationship still holds (but with shifts in the ranks of the core journals). Therefore, the *core* of science and scholarship is relatively small in terms of numbers of *journals*. Numbers of *articles* is another matter. These same 152 core journals published only about 25% of all the articles — proof that we can and must be more selective.

Bradford's distribution tells us something about journal publication and inter-journal citations. But my own law of concentration shows that the same core journals dominate a large number of seemingly separate fields.⁶ Figure 1 is a matrix illustrating this point: the same group of journals turns up in a variety of medical and scientific specialties. The blank spaces indicate particular specialty journals of importance to only one field.

ISI's *Journal Citation Reports*[™] (*JCR*[™]) shows that in the past decade the greatest growth in the literature was due not so much to *new* journals as to an increase in the number of articles in existing journals. Many new journals were simply expansions of existing ones: *The Journal of the Chemical Society* (London) was divided into six separate sections, and the American Chemical Society journals spawned new titles like *Biochemistry*, which has quickly become a core journal.

What if citation frequency were made the sole criterion of selection? Massive weeding would have to follow. Out of the millions of journal articles that could be cited, one-third to one-half will *never* be cited. Never. These arti-

cles may be read, but not cited. In fact, according to Halbert and Ackoff the average article is read by less than five people after it is published.⁷ Every librarian knows that some journal issues have never been touched by a single reader.

Assume for a moment that the world's literature of books and journal articles totals about 10 million items. If you prefer, make it 20 million items. My point is that less than one percent of all this material will ever be cited frequently enough to command the attention of one familiar with citation frequencies. Thus only 100,000 books and articles — or at the most 200,000 — can form the active core of a library able to provide copies of 90% of all future citations.

This "rational" approach to journal collection can be taken a step further. Instead of using journals in their present form and arrangement, we might collect only reprints of highly cited articles. These would have the highest likelihood of being used again. Derek Price has proposed a similar idea:⁸ publishing a journal consisting entirely of heavily cited papers called the *Journal of Really Important Papers*.

The idea is not so farfetched. In fact, the cluster data that ISI is now compiling as part of its *JCR* will greatly facilitate reprint collections of this kind. A group of libraries might get together to form a purchasing cooperative that could finance such reprinting on a large scale. I can imagine a collection of just 2,000 volumes, each containing 25 articles, that would satisfy 80 to 90% of a library's journal reference needs!⁹

Naturally, these initial 2,000 volumes would have to be augmented each year by about 5,000 newly identified items (about 200 volumes), both new and old, that had reached a specified citation threshold. But at the same time, many of the articles in the original collection will fall *below* the citation threshold, and would be discarded.

Such a core collection of heavily-cited articles would comprise a true no-growth library. However, I suspect that most libraries would choose to store the heavily-used articles even after they fall below the citation threshold. Oh well! At least the library would now grow *arithmetically* rather than *geometrically*. Probably most libraries would compromise, and settle for linear growth — and would again periodically face the weeding problem.

ISI has already begun the research necessary to identify, collect, publish and market a core collection of heavily-cited articles. Our cost estimates will be based on two important assumptions: that ISI would properly compensate publishers for use of their copyrighted materials, and that librarians providing hard-copy service of copyrighted materials would also compensate publishers.

In closing, I should like to call your attention to a letter published over 3 years ago:¹⁰

BRADFORD'S LAW

Sir, — In considering the application of Bradford's Law of Dispersion (Fairthorne A. J. *Doc.* 25:319, 1969) as a guide to acquisition policy in the research library or information centre it is pleasant to contemplate a bibliophilic Utopia of a complete collection in a library with unlimited space and acquisition funds. Utopias are rarely found, however, and the library *does* have limited resources. Given this restriction, the librarian or acquisitions specialist, in even the largest and most pecunious libraries, must make choices. These choices are rational only to the extent that the library collection maximises the timely provision of requested documents to the satisfaction of the largest number.

In this light, A. Faser's letter (*Nature* 227:101, 1970) suggesting that a library is derelict in not purchasing a specialized journal of interest to only one user treats the occasional request with the same degree of importance as the ongoing demand for the heavily used journals. An inventory policy in a department or food store, parts-supply depot, manufacturing concern or library, based on ignoring frequency-of-demand distributions, leads to inefficient allocation of resources. Designers of sewer and flood control systems know they cannot design economic drainpipe and culvert systems of sufficient capacity to handle the runoff from the one-in-a-thousand chance that rainfall will exceed, say, 6 inches in any 1 h period.

And mass merchandisers stock only a few or no items in the extremely low and high size ranges of shoes, hats and all attire in between.

Bradford's Law promulgates that a library can supply *most* of the requests for material with a *relatively modest inventory* of book and journal titles, geared to the *normal* pattern of demand. This demand pattern is one in which a relatively few items from among all possible items in the inventory satisfy a majority of the actual transactions. Progressively fewer transactions are satisfied from the balance of the inventory, or from further augmentation of the number of titles held. Abiding by the Bradford distribution, then, is an important factor in the library's overall success at demand-fulfilment.

The most efficient way for a library to exploit its collection and maximize utilization of its document file is to share its bibliographic resources with as many patrons as possible. It cannot *reasonably* be expected to serve *every* individual request. Carried to the extreme, if the only requests were one-time requests, there could not be an *economic* central library. The most efficient way of handling such a situation would be for each individual to have his own private collection.

Yours faithfully,

Melvin Weinstock

Institute for Scientific Information
Philadelphia, Pa. 19106

Many scientists assume that librarians make no judgments whatsoever in journal selection — that they are a captive market. It is true that many librarians have been guilty of poor administration of their journal collection — but it is true also that they have had only a minimum of hard data to work with. Fortunately, it is now possible to distinguish among books and journals — with accuracy, precision, and objectivity — and thus it is possible to make more scientific decisions.

1. Paper presented at the Associated Colleges of the Midwest Conference on Space, Growth and Performance Problems of College and University Libraries, Chicago, 17-18 April 1975.
2. Gore D. Zero growth for the college library. *College Management* 9:12-14, August/September 1974.
3. Williams J F & Pings V M. *A study of the access to the scholarly record from a hospital health science core collection*. Library and Information Service Center Reports. (Detroit: Wayne State University School of Medicine Report No. 54, January 1970).
4. Bradford S C. *Documentation*. (Washington, D.C.: Public Affairs Press, 1950), 156 pp.
5. Garfield E. Citation analysis as a tool in journal evaluation. *Science* 178:471-79, 1972. Reprinted in: *Current Contents*®(CC®) No. 6, 7 February 1973, p. 7-24.
6. ———. The mystery of the transposed journal lists; wherein Bradford's law of scattering is generalized according to Garfield's law of concentration. *CC*No. 31, 4 August 1971, p. 5-6.
7. Halbert M H & Ackoff R L. An operations research study of the dissemination of scientific information. In: *Proceedings of the International Conference on Scientific Information, Washington, D.C., November 16-21, 1958*. 2 vols., 1637 pp. (Washington, D.C.: National Academy of Sciences — National Research Council, 1959), vol. 1, p. 97-130.
8. Price D J D. Networks of scientific papers. *Science* 149:510-15, 1965.
9. Line M B, Sandison A & MacGregor J. *Patterns of citations to articles within journals; a preliminary test of scatter, concentration and obsolescence*. Bath University Reports. (Bath: Bath University Library Report No. 2, October 1972), 33 pp., ISBN 0 900843 27 6.
10. Weinstock M. Bradford's law. [Letter to the editor of] *Nature* 233:434, 1971.