



# Lies, Damned Lies, ... and ARL Statistics?

*by Kendon Stubbs, University of Virginia*

April 8, 1986

## I

ARL washes its dirty linen in committees. It is convenient to begin with this metaphorical statement about the recent work of the Committee on ARL Statistics. To have linen to wash is something. The publication of *The Gerould Statistics* demonstrates that ARL is the custodian of a statistical series stretching back nearly eighty years -- one of the oldest continuous annual series of library data in the world. To assert that this linen does not need washing from time to time would be rodomontade rightly disdained by anyone familiar with the vagaries of data compilations.

For several years the Statistics Committee has been examining the blemishes in the ARL data. This paper is intended to summarize the committee's work under the topics of (1) problems with the data that ARL collects; (2) problems with what ARL does not collect; (3) problems with data collection generally; and (4) proposed courses of action by the committee.

## II

*The ARL Statistics as a whole are a very unsatisfactory, dirty source of data. --Robert Hayes*

In 1914 Phineas Windsor, the Librarian of the University of Illinois, sent his president a copy of the latest statistics of university libraries, just issued by James T. Gerould at Minnesota. Mr. Windsor no doubt wanted to impress on the president that not only was Illinois still smaller than Oberlin in the size of its collections, but it even remained behind upstart libraries like Berkeley. President James, however, took a different view of the matter in his reply the next day: "I noted that while Illinois only added 29,000 volumes during the year 1913-14 and Yale added 37,000, we spent for the 29,000 \$86,000, while Yale spent only \$34,000. ... Is Yale getting much better bargains, or are we buying more expensive books, or what?" And the chilling conclusion, "I should like to have you analyze these things and bring them into harmony with the proposed budget of next year and explain the same."

Seventy years later the chief officer of another university opened *The Chronicle of Higher Education* to find the latest ARL statistics. After a little study he called in the library director and said: You spent 4.1 million dollars and added 120,000 volumes, while this other library spent 4.9 million and added 240,000. Did you pay nearly twice as much per volume? Are we buying more expensive books, or what?

The ratio of materials expenditures to added volumes is of course not a very useful measure of anything. Added volumes are one measure of cataloging productivity, and materials expenditures may cover a wide range, from monographs and serials to OCLC or RLG charges. Nevertheless, the question that university presidents have been posing for seventy years points to some persistent problems in the ARL statistics.

In William Kruskal's terms, some of the problems in the statistics are problems of smoothness: numbers that are erratic well beyond the bounds of typical fluctuations in the ARL data. For example, there was the library that reported a few years ago that it had added 300,000 volumes -- 112,000 more than the Library of Congress. Another library had 90 professionals in 1983, 68 in 1984, and 97 in 1985. Kruskal notes that problems of smoothness could be detected even by a visitor from outer space.

Transgressions of general knowledge, which are visible to those who know a little bit about how research libraries work, represent a different kind of problem. An example is the library that one year ranked in the top twenty in added volumes and in the bottom ten in staff, leading one to wonder at least about the library's secret of staff efficiency. Another example (which also may be considered exaggerated smoothness) is the library that for six years has reported exactly the same number of current serials.

A third kind of problem is due to the ambiguities of classification. At least one library has been counting microforms with volumes held and added volumes. Some libraries pay for bibliographic utilities or literature searching from book funds. The mysterious category of other operating expenditures seems to include a little of everything, probably even the library kitchen sink. Closely examined, a category like materials expenditures seems only a shade less fanciful than classifications noted by Borges. He relates, for example, that class 179 of the Bibliographical Institute of Brussels included books on "Cruelty to animals. Protection of animals. Moral implications of duelling and suicide. Various vices and defects. Various virtues and qualities."

The Committee on ARL Statistics has recently been taking a close look at smoothness, transgressions of general knowledge, and ambiguities of classifications. As a starting-point, the committee has been studying three measures of resources -- volumes held, gross volumes added, and current serials -- and two measures of expenditures -- expenditures for library materials and expenditures for serials. Here are some of the problems that the committee has been uncovering.

Volumes held: Some libraries include all of their government documents in volumes held; some include part of their documents; some included none. One library reported 2.8 million volumes, including all its documents; another reported 2.7 million, excluding 130,000 documents. If ARL members systematically included or excluded documents, there would be a radical shifting of their rankings with respect to volumes held.

Lesser problems are that a dozen members are counting holdings in bibliographical units rather than physical volumes, and some libraries include microforms with volumes held.

Gross volumes added: This is a higgledy-piggledy combination of purchased monographs and serials, gifts, documents, volumes from omnipresent backlogs, microforms and sheet maps in at least one library, and other materials too numerous to mention. According to the ANSI definition, added volumes are physical units that have been cataloged, classified, and made ready for use. There may be, however, 118 different opinions in ARL about the exact meaning of "cataloged, classified, and made ready for use."

Current serials: In 1984, 75 ARL members counted documents among current serials; 30 omitted documents from serials. The 75 that included documents paid an average of \$55.62 per serial. The 30 that excluded documents paid \$66.89. The president of one of the 30 universities might think that his library was buying more expensive journals than one in the 75 libraries. In fact, the serial counts in the 75 libraries may be inflated by several thousand "free" government documents.

Expenditures for library materials: This includes expenditures mainly for monographs and serials,

but also microforms, AV materials, maps, manuscripts, and so on. About half a dozen members seem to be including bibliographic utility charges in this category. Others, in unknown numbers, pay for literature searching or binding from the materials fund.

Expenditures for serials: The problem with this category is that ARL does not have figures on how many serials are purchased. The category of current serials include both purchases and gifts, and may also be a little fuzzy at the edges in areas such as monographic serials.

### III

*We need to do something about our statistics if they are to be used for decisions. We are finding our in-house management and statistics diverge more and more from the ARL data elements. --Patricia Battin*

If ARL is open to criticism for the data it collects, it is also being reproached for what it does not collect. A university president has written, "For one, I would de-emphasize currently quantitative measure of library status such as the annual ARL statistical rankings. . . . To judge the quality of libraries, for example, only by the number of volumes currently held or the number added each year may encourage the very competitive behavior which a commitment to interinstitutional cooperation and complementary collection development would deter. . . . Might there be some way in which the ARL surveys would include -- even in their quantitative data -- some measures of success in reducing duplication within and between institutions or improving preservation of existing materials?" It is true that the ARL statistics represent the traditional way of looking at libraries in terms of their resources in collections, staffing, and expenditures: the library as dragon guarding its goldhoard and content merely to gather up more treasure. Until recently, ARL has not paid much attention to measuring whether or how the treasure is being used -- in other words, to what in library jargon is called performance or access or service.

For some years the Statistics Committee has been looking at new ways of measuring research libraries. One outcome was Paul Kantor's manual on objective performance measures. Another is the supplementary statistics collected last year and this year. A main object of the supplementary statistics was to see how hard it is for ARL libraries to present new statistics in areas such as access and staff organization. It turned out that 8% of ARL members did not know (or at least could not easily explain) how many items they circulated; 16% did not know how many monographs they purchased in 1984; 23% did not have data on reference activities; 4% were unable to say how many branch libraries they administered. When nearly one of every ten libraries has trouble with a traditional question like number of circulations, the committee has been in a quandary about more probing measures of access.

In order to elicit reactions from the membership, the committee sent out a questionnaire earlier this year. In response to the question "what additional data elements, new compilations or data, or access measures would you like to see?" members named 78 additional elements of compilations they would support. Among those were success rates of reference transactions; expenditures for document delivery; gift titles cataloged per year; impact of bibliographic instruction; number of people leaving the library annually; "what is spent to maintain a square foot;" size of uncataloged backlogs; turnaround time for ILL; cost studies of acquisitions, reserve, binding, preservation, original cataloging vs. CIP or copy cataloging; and so on. The highest level of agreement among respondents was 35% in favor of an ARL price index. 15% favored data on online services, "access measures," and costs of automation. Most other of the 78 suggestions were made by 5% or fewer of members replying.

In thinking about these 78 suggestions, the committee has been acutely aware of the costs of collecting data, especially good data. Look at studies by accomplished researchers such as Paul

Kantor and Malcolm Getz, and you can see the painstaking care and time required for data that are smooth, do not transgress general knowledge, and are not subject to classification ambiguities. There are other lesser-known costs of new statistics. One of the best measures of both access and performance is the measurement of availability in Kantor's Objective Performance Measures. During the development of this manual, five ARL libraries tried out the measurement of availability, which computes the percentage of items immediately available to users when they come to the library to seek them. The percentages in the sample ranged from 42% to 59%, with a median of 48%; meaning that in these research libraries users can quickly put their hands on only one of every two items they want. Publication of this kind of data for 118 ARL members might be salutary from some standpoints, but it could also arouse questions as intense as current questioning of added volumes and materials expenditures.

The committee therefore continues to study new data elements that may be worth the cost to ARL members. Should the Association test the slippery slopes of access; should it begin to measure literature searching or automated services? These are questions the committee's current schedule.

#### IV

*None of us really understands what's going on with all these numbers.* -- David Stockman

If the ARL statistics are unsatisfactory and dirty, and anyway are not measuring significant aspects of research libraries, it is some comfort to find that no one else is doing much better; and some very respectable organizations are doing worse. Kruskal proposed that "A reasonably perceptive person, with some common sense and a head for figures, can sit down with almost any structured and substantial data set or statistical compilation and find strangelooking numbers in less than an hour."

A famous example of strange figures is known to statisticians as the Case of the Teen-age Widows. The 1950 U.S. Census reported -- erroneously, as it turned out -- 565 14-year-old widows and 1,670 14-year-old widowers. The Census is mandated by the Constitution; immense time and labor are devoted to it; and as a result it is one of the most sophisticated of all statistical compilations. Yet each Census offers new tidbits of piquant or exotic information.

In 1960 the Census revealed that 62 females, aged 15 through 19 years, had 12 or more children. In 1970, 2,983 14-year-old males were widowers; and 289 14-year-old females had been both widowed and divorced. The 1980 Census of Occupations informs us that 183,439 people called themselves librarians. 85% of these, not unexpectedly, worked in elementary and secondary schools, colleges and universities, and public and private libraries. Of the remaining 15%, however, 17 people working in gas stations considered their occupation librarianship. There were also 22 librarians at automobile dealers, 5 in optometrists' offices, 19 in trucking, 6 in "sugar and confectionary products," 22 in fishing, trapping, and hunting, and 1 in a funeral home.

Lack of smoothness, transgressions of general knowledge, and ambiguous classifications also infect educational statistics, including the standard national data collections. The Association for Institutional Research -- composed of the professionals who count whatever is countable in colleges and universities -- periodically wrings its hands over the messy numbers that are its lot in life. Ask your institutional research office what the FTE enrollment is on your campus, and you may get one answer on Monday and a different one on Tuesday. 50% of colleges and universities define a fulltime course load as 15 credit hours; 30% define it as 12 hours; others prefer anything from 11 to 16.5 hours. One institution uses six different methods of calculating FTE's for different reports throughout the year. Some universities count non-credit courses as hours of instruction; others as hours of public service. Some universities include technical-vocational students in their enrollments; others exclude them. Some include them in one report but not in others. The 1980

Census discovered 33% more college students in one state than the state had reported on its Office of Education HEGIS forms.

Financial data leave something to be desired. NACUBO has been moved to caution against national compilations. NACUBO's report found that expenditures of "medical schools, central administration, extension and research centers, institutes, and programs" are reported by some institutions, not by others. In at least 13 states tuition and fees are reappropriated by legislatures and thus counted by the universities under both tuition and state appropriations.

National library data compilations have been an easy target for years and used to elicit scorn and indignation in library journals; but apparently writers have grown weary of those pastimes. Mary Jo Lynch of the ALA Office of Research, together with her Advisory Committee on National Library Data Collection Project, spent several years convincing the U.S. National Center for Education Statistics to adopt a rational scheme of data categories for school, public, and academic libraries. But NCES has passed on to the graveyard of government acronyms, and its successor needs all our encouragement as it pursues the heroic scheme.

There are many more examples, but we are undoubtedly far enough along the downward path to morose delectation.

## V

*Statistics, to be sure, are not always reliable, but we have nothing better, and we must make as much of them as we can.*

-- Charles Angoff and H.L. Mencken

A starting-point for making as much as we can of the ARL statistics is to note that it is only a small proportion of data values that exhibit problems. The Statistics Committee judges that anomalies are visible in less than 5% of the ARL figures: about the same percentage that one would expect in the best of all possible statistical worlds. As a result, statistical analyses of the ARL data can produce remarkable reliable results. For example, regression analysis of the 1985 data shows that ARL libraries employ 19 professionals plus 1 for every 43,000 volumes held. This model is more accurate than models typically used for decision-making in many areas of psychology, sociology, economics, and even the physical sciences.

But alas, the university president reading the Chronicle is not likely to be looking for valid averages and well-behaved residuals and high coefficients of determination. He is likely to be putting his finger on the figures that demonstrate that a peer is paying only half as much per volume as his own library. Presidents (and library directors) have to keep in mind the story of the trout fisher who drowned in a stream that was only six inches deep . . . on the average.

The Statistics Committee therefore believes that it is worthwhile to wash ARL's statistical dirty linen. The committee agrees with the comment of one director in the recent survey: "ARL should focus its attention on improving its current statistics-gathering mechanisms before undertaking new initiatives. That does not mean we do not need additional data -- only that the quality of the data currently collected needs to be improved." The committee intends to concentrate on five data categories: volumes held, gross volumes added, current serials, expenditures for library materials, expenditures for serials. The committee proposes a survey of members to determine what is being counted in each of these categories. With the results of this survey the committee can better judge how to count or not count government documents, how to sort out the different ingredients in the materials expenditure bouillabaisse, and so on. At the same time the committee will continue to explore new data categories where utility may justify costs of collection.

In this matter the committee and the members are revisiting old battlefields. Forty years ago, at its 24th annual meeting, ARL accepted a proposal by Robert Downs that the members should count their holdings in bibliographical units, rather than physical volumes. Two years later, in 1947, a survey of members showed that half were counting in bibliographical units and half in physical volumes. In 1948, ARL appointed a new Committee on Counting Library Holdings, chaired by Guy Lyle. In 1949, at the 33rd annual meeting, the Lyle committee recommended that holdings be counted in physical volumes, rather than bibliographical units. Speaking to the motion, Downs said "that he had once thought uniformity possible but that he had become disillusioned on this subject and believed that no action taken here would have much effect." The members thereupon voted to record in the annual statistics whether counts were in physical volumes or bibliographical units; but no further action was taken on the Lyle proposal. By the 1960 winter meeting "the urgent need for uniform policies in counting library holdings was stressed by Jens Nyholm" But the meeting ended without any plan to meet the urgent need.

The current Statistics Committee believes that the ARL data can be redeemed. We urge the members to unite in this long-delayed endeavor.



[ARL Statistics and Measurement Program Home](#) | [ARL Home](#)

---

Maintained by [ARL Web Administrator](#)

© Association of Research Libraries, Washington, DC

*Last updated: Thursday 8 August, 2002.*