

Management Data for Collection Analysis and Development

BY DAVID A. KRONICK, PH.D., *Director*
VIRGINIA M. BOWDEN, *Associate Director*

The Library
The University of Texas Health Science Center at San Antonio
San Antonio, Texas

ABSTRACT

Sound management data are needed to evaluate the collections of health sciences libraries. This study reports the utilization of computer data bases to compare the library collections of The University of Texas Health Science Center at San Antonio, The University of Texas Medical Branch, and the National Library of Medicine's CATLINE data base. The imprint dates of the records of two libraries are compared to measure acquisitions rates. Subject profiles for the Q and W classes demonstrate the similarity of the collections. Reasons for the variances are considered.

THE library collection forms the basis from which most library services are derived. The selection and acquisition of library materials are, therefore, among the most important of library activities. However, as one reads the literature, one becomes aware that selection is one of the least precise and least understood of library processes. Despite the fact that there have been many efforts to measure and evaluate library collections [1-3], many of the measures are regarded as not very meaningful and many questions are left unanswered. For example, there are few data concerning the relationship of the size of a collection to its currency and quality. How do the collections of libraries serving similar programs compare? At what rates and frequencies do different libraries acquire material? How adequately is the literature in any particular subject being covered?

Answers to these questions are becoming more important as the rates of publication and the costs of acquiring and storing library materials continue to increase. In 1960-1961 the average library budget for academic health sciences centers was \$57,471; by 1973-1974 the average had risen to \$328,093 [4]; and only two years later, in 1975-1976, it was \$524,102 [5]. One way in which costs may be controlled is through the development of cooperative acquisitions programs in which institutions complement each other's

collections rather than duplicate them. In order to accomplish this objective sound management data, upon which cooperative acquisitions programs can be based, must be acquired.

The steadily expanding use of computers in library operations and the increasing availability of library records in machine-readable form have made the collection and analysis of the kind of data needed for managing collection activities much more feasible. This article reports an attempt to explore the relationship between health sciences library collections by using machine-readable cataloging data to obtain management information that might be useful in planning a cooperative acquisitions program for libraries in Region IX (TALON) of the Regional Medical Library Program.

DATA COLLECTION

The collections studied were those of The University of Texas Medical Branch at Galveston (UTMB) and The University of Texas Health Science Center at San Antonio (UTHSCSA). The data used had been gathered in the process of producing a union catalog of monographs for the TALON region [6]. Of the twelve libraries with data in the TALON catalog, UTMB and UTHSCSA were selected for this pilot project because the union catalog data base included the majority of their holdings: 32,257 records for UTMB and 41,725 for UTHSCSA. Both institutions are reasonably comparable in programs and support, except that UTHSCSA also serves a dental school. However, there is a vast discrepancy in terms of age because UTMB is one of the oldest libraries in the state of Texas, having its origins in 1891, whereas UTHSCSA accepted its first students in 1968.

Numeric breakdowns of the UTMB and UTHSCSA data were generated by a computer program. The number of records in specific call letter groupings were listed for each of the years between 1966 and 1976, with 1965 and earlier im-

prints included in a single category. This program was run in 1976 against the UTHSCSA data and again in 1977. The UTMB data were available for analysis in 1977 only.

The National Library of Medicine's (NLM) cataloging was studied through the use of CATLINE, an NLM on-line file that contains cataloging data for all monographs and serials cataloged by NLM since 1965. On-line retrieval techniques described by Byrd [7] were used to find the number of postings for each class.

ACQUISITIONS RATE

To measure the acquisitions rate the imprint dates of the records added to two data bases were compared. The two UTHSCSA computer analyses generated in 1976 and 1977 provided a measure of the imprint dates added by UTHSCSA in a one-year period. The NLM data were obtained by querying CATLINE for record input dates greater than December 1975 but less than January 1977. The authors recognize that input date does not represent the date the materials were acquired, because books may remain in a processing backlog for some time. However, the input date does provide some measure of the timeliness with which bibliographic records are made available and, therefore, the speed with which materials are made available to the user.

Table 1 displays the collection growth for the UTHSCSA and CATLINE data bases arranged by imprint year. Materials with no imprint year, such as open entries, are included with the pre-1966 group. The table shows the number of

records in the data base with the specified imprint date, both at the beginning and at the end of the one-year period. UTHSCSA had 1,907 records with an imprint of 1967 in its data base on January 1, 1976 (column 1), and 1,946 on January 1, 1977 (column 2), and it added 39 records with 1967 imprints (column 3), for a gain of 2.05% for imprints of that year. Columns 5 through 8 show similar data for the CATLINE data base. The table also indicates that 19.41% of the records input for UTHSCSA bear the imprints of the current year (1976), 889 out of a total of 4,580, while the corresponding figure for CATLINE is 33.95%. This indicates a significant acquisitions lag for UTHSCSA and can be considered both a reflection of and a justification for the use of centralized cataloging. Otherwise, the totals by imprint years (columns 2 and 6) tend to show some uniformity in growth over the years, that is, that CATLINE tends to add between 10,000 and 14,000 records for any imprint year, as against 2,000 to 2,500 for UTHSCSA. The table also shows that the data tend to stabilize—drop below 10%—for materials with imprints three years or older.

The data tend to be distorted to some extent by the fact that UTHSCSA represents a young collection as compared to the collection represented by CATLINE, because NLM has been collecting materials for well over a century. UTHSCSA would tend, therefore, to add older material that was already in NLM's collection by 1965 and, therefore, not in CATLINE. This is made clear when the data in Table 1 are reformatted to show the cumulative gains as percentages of the total number of records added

TABLE 1
COLLECTION GROWTH—ONE-YEAR CHANGE, BY IMPRINT DATE

Imprint year	UTHSCSA				NLM			
	Jan. 1976	Jan. 1977	Gain	% Gain	Jan. 1976	Jan. 1977	Gain	% Gain
	1	2	3	4	5	6	7	8
1966	1,519	1,554	35	2.30	10,270	10,383	113	1.10
1967	1,907	1,946	39	2.05	11,982	12,096	114	.95
1968	2,109	2,170	61	2.89	13,650	13,870	220	1.61
1969	2,236	2,284	48	2.15	14,122	14,317	195	1.38
1970	2,620	2,647	27	1.03	13,435	13,662	227	1.69
1971	2,503	2,551	48	1.92	11,005	11,314	309	2.81
1972	2,213	2,318	105	4.74	10,542	10,937	395	3.75
1973	2,075	2,233	158	7.61	11,228	11,818	590	5.25
1974	1,564	2,084	520	33.25	7,409	8,649	1,240	16.74
1975	351	1,817	1,466	417.66	4,111	8,427	4,316	104.99
1976	—	889	889	—	—	4,240	4,240	—
Total	39,018	43,598	4,580	—	145,315	157,803	12,488	—

COLLECTION ANALYSIS AND DEVELOPMENT

TABLE 2
TITLES ADDED, BY IMPRINT YEAR

Year	CATLINE				UTHSCSA			
	No. added	Cum. total	% of total	% Cum.	No. added	Cum. total	% of total	% 1966-1967
	1	2	3	4	5	6	7	8
1966	113	113	0.90	0.90	35	35	0.76	0.76
1967	114	227	0.91	1.82	39	74	0.85	1.62
1968	220	447	1.76	3.58	61	135	1.33	2.95
1969	195	642	1.56	5.14	48	183	1.05	4.10
1970	227	869	1.82	6.96	27	210	0.59	4.59
1971	309	1,178	2.47	9.43	48	258	1.05	5.63
1972	395	1,573	3.16	12.60	105	363	2.29	7.93
1973	590	2,163	4.72	17.32	158	521	3.45	11.38
1974	1,240	3,403	9.93	27.25	520	1,041	11.35	22.73
1975	4,316	7,719	34.56	61.81	1,466	2,507	32.01	54.74
1976	4,240	11,959	33.95	95.76	889	3,396	19.41	74.15
Other	529	12,488	4.24	100.00	1,184	4,580	25.85	100.00
Total	12,488		100.00		4,580		100.00	

in the period (see Table 2). Thus, only 4.24% of the records added to CATLINE had either imprints before 1966 or no imprint, while that figure was 25.85% in the case of UTHSCSA (columns 3 and 7). The table also shows that a total of 78.44% of the records added to CATLINE had imprints in the last three years, 1974-1976, while the corresponding figure for UTHSCSA was 62.77%. One major factor for this is NLM's participation in the "Cataloging in Publication"* program. Another factor is that one of the chief selection sources for UTHSCSA is NLM's cataloging proof sheets.

When the percentage figures are adjusted to eliminate the pre-1966 imprints from the calculations, the proportions are quite different: 81.91% of the records added by CATLINE fall in the three-year period, while the corresponding figure for UTHSCSA is 84.66%. The data thus show that, although CATLINE adds more data for the current imprint year, UTHSCSA adds a proportionate amount when the input is averaged out over the last three years. The similarity between the rates of growth in the UTHSCSA and CATLINE data bases indicates that libraries, even large research libraries, continue to add a significant number of records much older than the current year. This suggests that valid studies of

duplication in acquisitions activities should start with imprints at least two, preferably three, years prior to the period being studied.

Imprint data can also provide some measure of the obsolescence of a collection, particularly when the entire shelf list is available in machine-readable form, as is the case with UTHSCSA. This factor is of particular importance in evaluating scientific research collections, because a collection of 100,000 volumes that has only 5% of its materials under ten years of age cannot be equated with a collection of similar scope that has 20% of its material in that age range. Table 1 shows that monographs with imprints in the ten-year period 1966-1975 totaled 19,097, or 48.94% of the monograph collection of UTHSCSA (column 1), and that at the end of the year materials ten years old or less (1967-1976) comprised 20,739, or 47.37% of the monograph collection (column 2).

SUBJECT PROFILES

Computer analysis of cataloging records may also be used to compare collections by subject coverage. Table 3 lists the number of records for the major classes for the CATLINE, UTHSCSA, and UTMB data bases. The CATLINE figures were obtained by querying CATLINE for postings by call number in either the "ØXNLM"† or

*This is really a misnomer; the cataloging actually occurs from proof, and thus prior to publication, and perhaps would better be called PPC for "Prepublication Cataloging."

†ØXNLM is the CATLINE tag for subject classifications for items that NLM puts in form classes, e.g., serials. Ø4NLM is the tag for NLM's own, or shelf list, classification.

KRONICK AND BOWDEN

TABLE 3
 HOLDINGS IN SPECIFIC CLASSES

Subject area	CATLINE		UTHSCSA		UTMB	
	Holdings	% Relative frequency	Holdings	% Relative frequency	Holdings	% Relative frequency
BF Psychology	3,143	2.97	620	1.63	680	2.34
Q Science	895	.85	287	.75	357	1.23
QA Mathematics	190	.18	211	.55	156	.54
QD Chemistry	1,304	1.23	608	1.60	332	1.14
QH Natural history	2,913	2.75	1,084	2.84	1,158	3.99
QK Botany	376	.36	73	.19	114	.39
QL Zoology	1,063	1.00	431	1.13	562	1.94
QP Physiology	167	.16	108	.28	127	.44
QS Anatomy	1,368	1.29	680	1.78	436	1.50
QT Physiology	2,628	2.48	628	1.65	525	1.81
QU Biochemistry	2,734	2.58	1,246	3.27	805	2.78
QV Pharmacology	4,754	4.49	1,477	3.88	1,202	4.14
QW Bacteriology-immunology	2,098	1.98	1,049	2.75	714	2.46
QX Parasitology	602	.57	130	.34	130	.45
QY Clinical pathology	1,250	1.18	506	1.33	297	1.02
QZ Pathology	2,298	2.17	1,084	2.84	758	2.61
SF Veterinary medicine	1,757	1.66	300	.79	126	.43
W Medical profession	4,774	4.51	1,943	5.10	1,409	4.86
WA Public health	6,909	6.53	1,376	3.61	926	3.19
WB Practice of medicine	5,141	4.86	1,766	4.63	1,045	3.60
WC Infectious diseases	1,877	1.77	735	1.93	547	1.89
WD Metabolic diseases, space medicine	1,268	1.20	371	.97	313	1.08
WE Musculoskeletal system	2,928	2.77	1,029	2.70	737	2.54
WF Respiratory system	1,692	1.60	535	1.40	377	1.30
WG Cardiovascular system	3,204	3.03	1,071	2.81	829	2.86
WH Hemic-lymphatic system	1,303	1.23	470	1.23	401	1.38
WI Gastrointestinal system	2,234	2.11	692	1.82	559	1.93
WJ Urogenital system	1,233	1.17	530	1.39	377	1.30
WK Endocrinology	1,416	1.34	584	1.53	447	1.54
WL Nervous system	4,070	3.85	1,719	4.51	1,455	5.02
WM Psychiatry	7,941	7.51	2,502	6.56	2,071	7.14
WN Radiology	1,895	1.79	776	2.04	408	1.41
WO Surgery	2,199	2.08	1,312	3.44	865	2.98
WP Gynecology	1,485	1.40	565	1.48	358	1.23
WQ Obstetrics	1,282	1.21	455	1.19	319	1.10
WR Dermatology	792	.75	285	.75	235	.81
WS Pediatrics	3,896	3.68	1,078	2.83	840	2.90
WT Geriatrics	1,029	.97	228	.60	208	.72
WU Dentistry	2,113	2.00	998	2.62	174	.60
WV Otorhinolaryngology	1,168	1.10	425	1.12	326	1.12
WW Ophthalmology	1,527	1.44	1,267	3.32	517	1.78
WX Hospitals	2,283	2.16	516	1.35	350	1.21
WY Nursing	2,937	2.78	1,070	2.81	983	3.39
WZ History of medicine	4,474	4.23	1,765	4.63	2,103	7.25
Z Bibliography-library science	3,172	3.00	1,527	4.01	1,348	4.65
Total	105,782		38,112		29,006	

COLLECTION ANALYSIS AND DEVELOPMENT

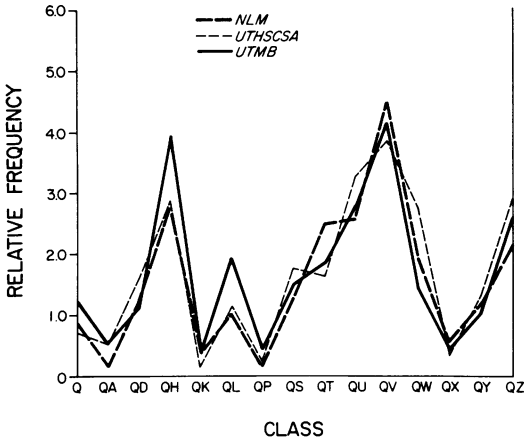


FIG. 1.—Subject profile of Q classes.

“NLM” field. Hence a work with a form call letter was not included unless an alternate call number was assigned. The UTMB and UTHSCSA data were generated by the same computer program that produced the imprint data.

To further explore relationships, we computed the relative frequencies of the classes listed in Table 3. These percentages were charted for the Q (science) classes in Fig. 1 and for the W (medicine) classes in Fig. 2. The two charts illustrate striking similarity among the subject profiles. We have considered possible reasons for some of the

differences. These, of course, would have to be tested on a title-by-title basis.

In the subject profiles (Fig. 1 and 2) there seems, strangely enough, to be a wider range of variation in the clinical sciences than there is in the basic sciences. Even in the clinical sciences, however, what is most noticeable is a tendency of the profiles to conform rather than to vary. Without a thorough analysis of the holdings of each of the libraries one can only speculate about the larger variations. The profile peaks in almost the same proportion for all three libraries in WM (psychiatry). This seems to indicate that the volume of purchasing in a library in a particular subject field may be directly related to its publishing volume (particularly high in the subject area of psychiatry.) In WA (public health), where there seems to be the greatest variation between NLM and the other two libraries, the explanation may be that NLM as a national library has a responsibility to collect public health materials in all the states, as well as from foreign governments, while the Texas institutions have responsibility primarily for their own state. Neither of these Texas schools has a formal program in public health, and the two therefore may not collect even on the research level, while NLM probably would wish to collect comprehensively. Another noticeable variation is in WU (dentistry). UTMB has no formal program in this area, while UTHSCSA does.

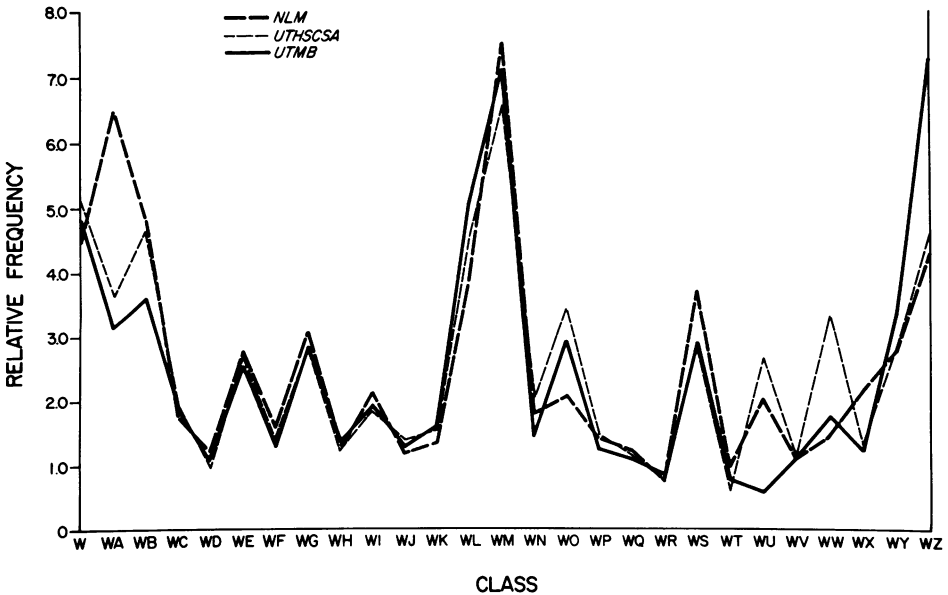


FIG. 2.—Subject profile of W classes.

Another comparison is possible using this same information extracted from the data bases: the number of records in each subject class in UTHSCSA as a percentage of those in CATLINE. Ignoring some of the anomalies like QA (mathematics), in which UTHSCSA has more records than CATLINE, and areas like WA (public health), regarding which we have already ventured an explanation, it is clear that CATLINE is stronger proportionately in BF (psychology), H (social sciences), QK (botany), and SF (veterinary medicine) than UTHSCSA. In the other areas the proportionate relationship between the collections of UTHSCSA and CATLINE seems to range between one-fourth and one-half, which seems to be reasonable for one between a regional and a national library. The reasons for the specific relationships, however, need further exploration.

In 1976 Byrd used CATLINE in a similar way to determine its subject coverage and compared it with a manual examination of the shelf lists of the University of South Dakota Health Science Library and the University of Nebraska Medical Center Library [7]. He compared them in a bar graph to which he added the data produced earlier by Fenske, who analyzed 8,137 books cataloged by the Washington University School of Medicine between 1965 and 1970 [8]. To compare these figures with the Byrd data we extrapolated the percentages for the four libraries he covered and added those for UTMB and UTHSCSA, to produce the data in Table 4. Despite significant variations in some of the broad classes, some of which Byrd attempts to explain (South Dakota for a two-year medical school before 1974), there is a remarkable consistency across the board.

SUMMARY AND CONCLUSIONS

The study of the UTHSCSA, UTMB, and CATLINE data bases is suggestive rather than definitive. Quantitative data have many limitations because the numbers may represent quite different titles. The figures for UTHSCSA are for volumes; those for UTMB are for titles. Because the ratio of volumes to titles is close and because they can be assumed to be evenly distributed, it was felt that a comparison would still be useful.

Although the relative frequencies of NLM's, UTMB's, and UTHSCSA's classes are strikingly similar, the CATLINE figures are usually two to four times greater than those for either UTMB or UTHSCSA. This is not surprising, because NLM is a national library with a mandate to collect comprehensively. Furthermore, NLM's policy is to not be selective by quality, whereas UTMB and UTHSCSA must be selective. Another area of disparity is foreign language material, as neither UTMB or UTHSCSA routinely add non-English monographs.

It is also reasonable to assume that some materials were classified in different ways. Jackson and Hendricks, in a study of the UTMB, UTHSCSA, and University of Texas Health Science Center at Dallas, found that approximately 12% of the titles held by at least two of the three libraries were assigned a different classification in each library [9]. NLM classifies its serials, technical reports, and some other material in form classes. Although the study used the alternate call number assigned, it is possible that these were not always assigned.

Comparisons of academic health sciences library collection development have been largely in

TABLE 4
SUBJECT PROFILE OF SIX LIBRARIES, BY MAJOR LIBRARY OF CONGRESS AND NLM CLASSES

Class	% of collection					
	NLM†	USD*†	NEB*†	WAU*†	UTMB	UTHSCSA
	1	2	3	4	5	6
A-P	9	2	11	6	12	11
Q	4	9	6	9	9	7
S-V	2	2	1	2	2	3
QS-QZ	13	26	17	17	15	16
W	65	59	60	59	57	59
Z	7	4	5	7	5	4
Total	100	100	100	100	100	100

*USD: University of South Dakota; NEB: University of Nebraska; WAU: Washington University, St. Louis.

†Extrapolated from Byrd [7].

terms of size, quantitative growth, and funds expended. This study has demonstrated that machine-readable data can be used to add two other important aspects: collection currency and subject coverage. The objectives of a cooperative acquisitions program vary from those of centralized processing centers. In the case of the latter the system is more efficient and economies are realized when duplication in acquiring materials is maximized. In cooperative acquisitions, on the other hand, the goal is to minimize duplication so as to use limited funds to broaden the resources available to the participating libraries. In either case it is important to be able to compare the rate and lag time among the participating libraries in acquiring materials. Jackson and Hendricks called attention to this problem in their study of the feasibility of centralized processing [9]. During the six-month period that they studied, only 800, or 9%, of 8,130 cards submitted to a union catalog represented duplications. They estimated that in the long range duplication among the five libraries studied might reach as high as 70%. Our study has not addressed itself at this time to the problem of duplication, but we feel we have demonstrated that it is likely to occur at any time in the acquisitions cycle, and that in studying duplication rates it is better to emphasize imprints two years old or older.

Another important consideration in developing cooperative acquisitions programs is the subject scope of the participating libraries. If the libraries tend to differ from each other in this regard, they will tend to purchase different books and to supplement each others' collections. A cooperative acquisitions program can then be based on making commitments to acquire materials in those subjects by which their collections are already characterized. However, if the libraries tend to be similar in subject scope they will tend to buy the same books unless an attempt is made to assign commitments for specific subjects, a difficult solution for libraries serving similar programs. It seems clear, therefore, that cooperative acquisi-

tions between institutions with similar academic and research missions need to be based on factors other than subject coverage.

ACKNOWLEDGMENTS

The authors wish to thank the Moody Medical Library of The University of Texas Medical Branch for the use of the UTMB collection data and for reviewing its accuracy. We also wish to thank Richard Day of the UTHSCSA Library for his assistance in searching CATLINE.

REFERENCES

1. LANCASTER, FREDERICK WILFRID. *The Measurement and Evaluation of Library Services*. Washington, D.C., Information Resources Press, 1977. p. 165-206.
2. BONN, G. S. Evaluation of the collection. *Libr. Trends* 22:265-304, Jan. 1974.
3. MACHLUP, F. Our libraries: can we measure their holdings and acquisitions? *AAUP Bull.* 62: 303-307, Oct. 1976.
4. CRAWFORD, S. Medical libraries in the United States 1960 through 1975. *JAMA* 237: 464-468, Jan. 31, 1977.
5. HENDRICKS, DONALD D., comp. *Medical Library Statistics 1975-1976*. Dallas, University of Texas Health Science Center at Dallas Library, 1977.
6. KRONICK, D. A., AND BOWDEN, V. M. A union catalog of monographs: another approach. *Bull. Med. Libr. Assoc.* 66: 281-289, July 1978.
7. BYRD, GARY D. The shelf list as an administrative and acquisitions tool. In: *Abstract Proceedings of Medical Library Association 75th Annual Meeting*, June 12-17, 1976, Minneapolis, Minnesota.
8. FENSKE, R. E. Mechanization of library procedures in the medium-sized medical library: XIV. Correlations between National Library of Medicine classification numbers and MeSH headings. *Bull. Med. Libr. Assoc.* 60:319-324, Apr. 1972.
9. JACKSON, EUGENE B., AND HENDRICKS, DONALD D. Feasibility of centralized processing for five medical library components of The University of Texas system. Austin, Texas, 1973. p. 55. Unpublished report.

Received May 8, 1978; accepted May 30, 1978.