

GLYN T. EVANS

Assistant Vice-Chancellor for Library Services
State University of New York
Central Administration

ALBERT BEILBY

Acting Director of Education and Training
Training Department
Binghamton Psychiatric Center

A Library Management Information System in a Multi-Campus Environment

Introduction

The Office of Library Services in the Central Administration of the State University of New York (SUNY) has, since 1975, been developing a library management information system based on the analysis of library and other bibliographic and academic data which are available in machine readable form. Although primarily designed for the SUNY libraries, the processes are applicable in other academic libraries because of the general availability of the data used in the system. The task has changed over the years as new ideas and opportunities were realized, as new appreciations of the obtained results were attained, and as the technical environment has evolved. Nonetheless, the fundamental structure of the system design has not changed since the first ideas in 1974.

This is an interim report. Progress has been agonizingly slow for two reasons. First, the difficulty of obtaining support and resources has been a real hindrance; the work has been squeezed into overcrowded schedules and ever-straitening budgets. Second, many of the machine-readable data which one confidently felt would be available in the late 1970s or very early 1980s are still not available. Some years, at least, will pass before the work can be completed *as we see it now*. Who knows what new ideas and opportunities will emerge as new results become available? Nonetheless, enough has been achieved to justify this report.

Environment

The State University of New York is a multi-campus university in New York State. It is composed of thirty-two state operated campuses and thirty-two community colleges which are administered by local county authorities. The senior colleges in New York City and ten community colleges in the city compose City University of New York (CUNY), a separate organization.

The total student head count enrolled at the state operated campuses (i.e., excluding community colleges) in 1982 was 385,000, with 29,200 faculty and staff. The total collection size is slightly more than 10.7 million cataloged volumes, growing at the rate of some 400,000 volumes per year. The acquisitions budget in 1982 was slightly more than \$12.8 million.

There are thirty-five separately administered and budgeted libraries within the SUNY state-operated campuses—the primary locus of these studies. They are composed of four university centers (including a law school library), four medical schools, twelve four-year colleges of arts and sciences, six two-year agricultural and technical colleges, four special colleges (forestry, maritime, optometry, and technology), and four statutory colleges (Alfred Ceramics, Cornell Agriculture and Human Ecology, Industrial Labor Relations and Veterinary Medicine libraries). All the variety one's heart could desire.

The Office of Library Services in the Central Administration of SUNY is charged with planning, developing and integrating the library resources of SUNY in support of its academic programs. In an early step to achieve this goal, the office contracted with OCLC in 1973 to provide services to State University and other New York participating libraries. As a result, the SUNY/OCLC Network is also administered by the Office of Library Services. The network now comprises 228 institutions (academic, public, school, law, medical, state agency, etc.) and for-profit institution libraries. A further 600 (approximately) are either sharing institutions or are members of processing centers or Regional Union Lists of Serials. Within the state, 70 percent of the independent higher educational institutions, 82 percent of public higher education, and 78 percent of the public library systems participate in the network.

As one of its services, the SUNY/OCLC Network stores and processes OCLC distribution tape records for its libraries at the SUNY Central Administration Computer Center. Currently over 10 million records are housed for the libraries, with the file growing by approximately 2 million each year. There is a clear relationship between this activity and the development of the library management information system.

It should be stressed that the Central Office of Library Services does not have responsibility for the direct operation of the campus libraries,

which report ultimately to each campus president. The central office has a planning function in which it tries to create as hospitable an environment as possible for the campus libraries. The office is guided by Stafford Beer's dictum that "the only feasible ultimate objective of systemic control is to hold the system within its natural boundaries."¹

Decision Support Systems

The computer has been seen as an aid to management from the time of its adoption by organizations as an administrative tool (as opposed to a research or production tool). It has been expected that data which emerged as a byproduct of (or could be coaxed from) a production operation would be collected, collated and analyzed to provide management with data which would improve the operation of the system. In library circles, circulation data seems to attract the most attention. It must be said, however, that recorded examples of such data actually being used to make decisions are extremely rare. A paper (such as that of F.H. Spaulding and R.O. Stanton from Bell Laboratories²) which records the effect circulation had on acquisition/selection decisions is a desert rose. Indeed, the development of management information systems seems to have created yet another battleground for internal control of any organization—well described by Peter Keen in his article "Information Systems and Organizational Change."³

The term and activity of management information systems is now becoming supplanted by decision support systems (DSS), the fundamental difference being that instead of providing passive displays of data (probably offline), a true decision support system goes one step further. The data are available online in a synthesized form from a variety of sources, and are presented to the administrator through a "friendly" terminal which supports modeling programs, color graphics and other facilities—all of which give the manager the opportunity to review the data and to test alternate strategies. King defines a decision support system as, "a computer-based system that the administrator uses to amplify or improve judgement. It is not a system that makes decisions."⁴ DSS software has the following capabilities: (1) report preparation and inquiry; (2) modeling language; (3) graphic displays; and (4) financial and statistical routines. Hopkins and Massy write:

The process of modelling is always one of synthesizing known facts, theories, and judgements into a meaningful pattern.

Models are about something; they purport to represent an aspect of something that exists, or might exist, in the real world. We call the object of a model the *reference system*. Thus a given reference system can in principle be represented by many different models, each one more-or-less

accurate with respect to certain characteristics of the system....Models need to be verified...[and] validated....Models are designed for a purpose.⁵

They also caution readers in a manner similar to King: "A good model—one that is simple but complete, stable yet adaptable—should make the quantifiable dimension of decision-making a far less mysterious place in which to operate, but it will not thereby lessen the burden of choice."⁶

The task at SUNY has been to build such a model, although in truth the work began before the term and the attractions of the DSS were developed. Some results have been obtained from the work to date, although the project is not yet complete. DSS systems hold the promise of reducing development time.

Stimuli

The SUNY Office of Library Services did not begin this work on an idle whim. The libraries of the university were under severe fiscal pressure—and we had little idea then that the economic conditions would continue to deteriorate as they have. Since other institutions have faced some of the pressures, it is useful to list those seen as critical at SUNY:

1. *Acquisition formula budgeting.* The university had adopted the Clapp-Jordan formula⁷ in 1968 and was using the formula to build the collection. In 1975, the formula was used by the Division of the Budget to cut back the acquisitions budgets for three of the university centers on the grounds that the libraries were, or soon would be "adequate" according to the formula. The loss of funds was two-thirds of \$1 million from the annual acquisition base. Further, a derivative of that formula was being promulgated by the State Education Department as a state-wide guideline.⁸ These actions were clearly not in the best interests of the university's libraries. The heat was on to develop another formula. As it happens, work had already begun on a discipline-based "formula," but was not complete in time to be offered as an alternative to the cuts in funds. (It is of interest that the results of a project by Evans, Beilby and Gifford completed in the process of developing an acquisitions formula, concluded that it is *not* possible to derive a "formula.")⁹ It is possible to develop an information system which will reflect the bibliographic components of the academic mission of a campus.
2. *Lack of an adequate statistical database.* Apart from gross budgeting data, and biennial Library General Information Survey (LIBGIS) reports,¹⁰ there was no firm database to describe the libraries, their collections, their successes, or their failures.

3. *Isolation of library data from academic administrative data.* Although library data were gathered from LIBGIS, and summarized from budget data, there was no link to relate them to academic program data such as the location and enrollment in courses and programs, level of programs (i.e., undergraduate or postgraduate) or location of programs. The chicken and egg question of the location of programs and location of library collections and which comes first could barely be asked, let alone answered. The one reliable element was dollars expended per full-time equivalent (FTE) student, achieved by dividing one number into another.
4. *Traditional library data emphasis on "inputs" rather than "outputs."* Apart from total circulation, and interlibrary loan (ILL) traffic, there was almost no emphasis on the collection of service data, which for a service organization is incomprehensible. (Reference statistics were added later.) LIBGIS surveys and the state guidelines emphasized space, facilities, staff, collection size, and collection growth, rather than performance. Not that those data are not important, they are. But one wants to know how successful, or useless, one's library is; its service, not its potential for service.) Furthermore, there was no evidence that the "official" requirements for statistical reports would change as machine-readable data became more available.
5. *The multi-campus environment.* Because of the multi-campus nature of the university, and because of the nature of the growing ties among libraries sharing the OCLC Network, it was clear from the beginning that any system design must accommodate that variety and that added dimension. Therefore, the system was designed from the ground up with multiple campuses and multiple academic programs and libraries in mind. Fortunately the program was prescient in that regard as fiscal crises have begun to force the "trades and affiliations" of academic programs within the university.
6. *Fiscal pressures.* The budgetary problems have already been cited. But there are other, subtle factors which should be drawn into account: the shift of monies to serials rather than monographs; new physical media; the growing necessity to purchase information-on-demand through database searches—or ILL—(as opposed to buying the potential to supply service through acquisitions); or the decision to retain or discard an item.
7. *Political pressures.* As would be expected, a sharp reduction in the acquisitions budget tended to attract attention to the problems with demands for a quick solution despite the absence of data.

A recitation of the earlier stimuli should not be interpreted as a criticism of the university or the profession. Rather it is criticism of the

conventional wisdom, and the reluctance to accept both the need for data, and the need to absorb them into operational decisions. It takes an unconscionable amount of time and effort to effect change.

Given these stimuli, and the growing availability of machine-readable data, it was decided to build a library management information system which would satisfy the following purposes:

1. to establish a model that would describe the acquisition/retention process in a multi-campus academic environment;
2. to develop and refine the available databases for inclusion in the model—i.e., to rely on data which have been acquired as the byproduct of a production operation;
3. to establish computer programs which would drive the model and provide reports;
4. to provide individual campus reports and system-wide reports—including time-series and trend reports; and
5. to integrate the reports into the planning processes of the libraries and the institutions.

The uses to which the management information would be put include:

1. justification of acquisition budgets;
2. support of the planning process, particularly among:
 - (a) academic programs and library programs,
 - (b) campuses, and
 - (c) campuses and their local disciplinary environments;
3. provision of specific campus/interdisciplinary reports by library disciplinary strengths and/or weaknesses;
4. support of the campus accreditation process; and
5. exploitation of the ability of the analytical programs to provide subject and/or form bibliographies by discipline for a campus, or group of campuses.

The Model

An early description of the model was reported in 1978,¹¹ but for the benefit of continuity it is briefly described here. The structure is that of the familiar five-box information system, comprising input, control, decision, output, and feedback (see fig. 1).

The central decisions in a library is the acquisition and retention decisions. The sum of these decisions is, in fact, the library. In an academic environment, the inputs into that decision are supply—i.e., what materials are available—and demand—i.e., what academic and research programs are supported by the library and for which community of users. The output

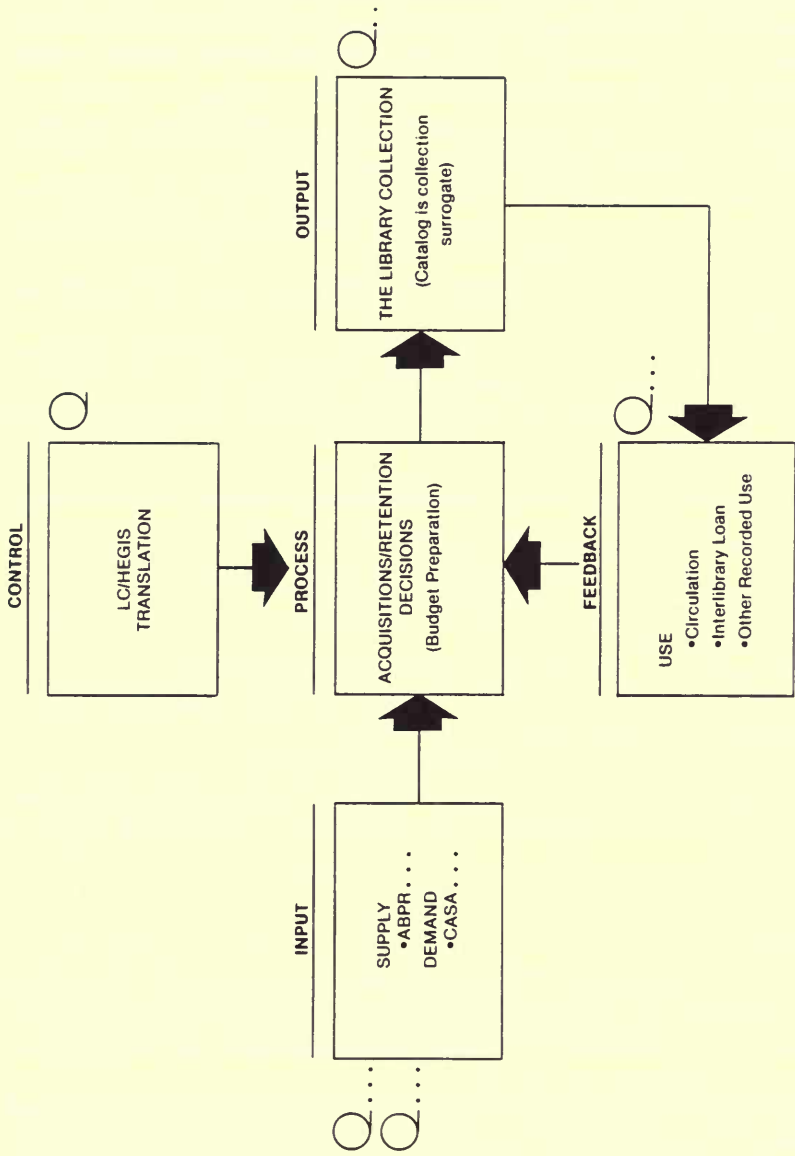


Fig. 1. Representation of the Five-Box Information System

of the decisions is the library, and the catalog is the surrogate of the library. The feedback into the decision (another source of input) is the use which is made of the collection, defined by the discipline and the community, and through the circulation, interlibrary loan and internal use. The control or program element is the software and tables which drive the management information system.

It was noted earlier that one of the criteria for the model design is that the machine-readable data to be used would be operational data. Such machine-readable data are available for each group described earlier although not at present in all SUNY libraries. The data are summarized as follows:

1. Input data (supply): Machine-readable files such as Bowker's Books-in-Print, or American Book Publishing Record (ABPR).
2. Input data (demand): Campus enrollment data described with the U.S. Office of Education, Higher Education General Information Survey (HEGIS) disciplinary codes.
3. Output data: OCLC distribution tapes containing local library versions and holdings of items cataloged into OCLC.
4. Feedback data: machine-readable files of circulation transactions from automated circulation systems; and machine-readable files of ILL transactions from online network ILL systems.

It is shown that with the exception of the Input Demand HEGIS Enrollment data, there is, or is likely to be, considerable consonance among the data files. They all are bibliographic data files, and will include, and transport, the same data elements among files depending upon the purpose of the transaction. All library files (Bowker, OCLC, circulation, and ILL) carry a Library of Congress (LC) classification number, and probably an LC card number and an ISBN. Three of the files carry, or can be made to carry, an OCLC number.

Files will also carry additional fixed field codes which assist in the selection of data for analyses, and transaction codes which define the nature of activity of the record itself. Examples of the latter are OCLC update, produce or cancel codes.

The data elements which are used for analyses (as opposed to the selection of records from a large file for analysis) in the segments of the programmed model are the OCLC number and LC class number, used singularly and in combination. Used by itself, the cooccurrence of the OCLC number is a measure of the degree of overlap among collections. The LC class number is an indicator of the subject strength of a library or a group of libraries. Using both elements in conjunction, it is possible to define both the collective subject strength of a group of libraries, and the

degree of uniqueness or commonality of holdings of a library or all libraries within each academic discipline.

As data become available following the installation of automated circulation systems use, databased on LC class number and OCLC number will be derived from ILL and circulation transaction files and entered into the decision box as feedback data.

All of the discussed elements are well understood by library and information professionals, but the academic administrative data used as the "demand" segment of the input component of the model are less familiar. The U.S. Office of Education's National Center for Educational Statistics requires an annual HEGIS survey, through which all institutions report the number of degrees awarded, the numbers and levels of students and faculty, etc. To facilitate reporting, a disciplinary taxonomy was established in 1971.¹² SUNY has developed an automated statistical reporting system which uses this taxonomy, plus fiscal information in a Course and Section Analysis (CASA) file, to produce annual statistical abstracts on trends and costs within the university.¹³ There is thus available a massive file of machine-readable data on the potential demand for library services by discipline and by the university community.

The immediate and obvious problem is that the HEGIS/CASA file does not carry any bibliographic data elements. However, the problem was overcome in the research project by Evans, Beilby and Gifford noted earlier which built conversion tables in which each term in the HEGIS taxonomy is expressed as a series of LC class numbers, creating, in effect, a series of mini-classifications. The ground rules are that (1) LC class numbers can be drawn from any part of the class schedules, and (2) LC class numbers may be used as many times as necessary. The HEGIS taxonomy has a two-tier structure in which major classes are divided into subclasses (see fig. 2). For its own statistical abstracts, SUNY has created a higher level amalgamation of classes designated as disciplines. There are ten such groups. The mechanism by which LC class terms can be assigned to a HEGIS subclass and subsequently amalgamated into higher levels is indicated in figure 3.

After the structure was defined, individual library subject specialists undertook to create the HEGIS/LC tables. It was found that over 13,000 LC classes were used to describe 494 HEGIS subclasses. (Figure 4 is a sample of entries taken from the African studies HEGIS class. The descriptions are taken from the LC class schedules.)

The use of LC class number, the HEGIS/LC tables, the OCLC record number and the campus code (OCLC's three-letter symbol) allows the identification of collection strengths and uniqueness related to teaching demands at any campus or within a group of campuses. Or conversely, their use can take a specific discipline or class and assess the relative campus strengths in that area. Since each record which is assigned to any

04 BIOLOGICAL SCIENCES

0400 Biological Science Unclassified
0401 Biology General
0402 Botany General
0403 Bacteriology
0404 Plant Pathology
0405 Plant Pharmacology
0406 Plant Physiology
0407 Zoology General
0408 Pathology, Human, Animal
0409 Pharmacology, Human, Animal
0410 Physiology, Human, Animal
0411 Microbiology
0412 Anatomy
0413 Histology
0414 Biochemistry
0415 Biophysics
0416 Molecular Biology
0417 Cell Biology, Cytology
:
:
:
0499

Fig. 2. Example of the HEGIS Subclass Structure

class is individually identified, it is possible to create subject bibliographies by class to be used both as reference tools and for accreditation assessment purposes.

The Process

In a simplified form, the following steps are taken to complete the analysis:

1. receive OCLC tape;
2. read and extract the selected record use for analysis;
3. process the extract tape in the analysis program by (a) matching the call with HEGIS/LC tables; (b) assigning it to levels (i.e., subclass, class discipline, institution); (c) counting; (d) matching with CASA

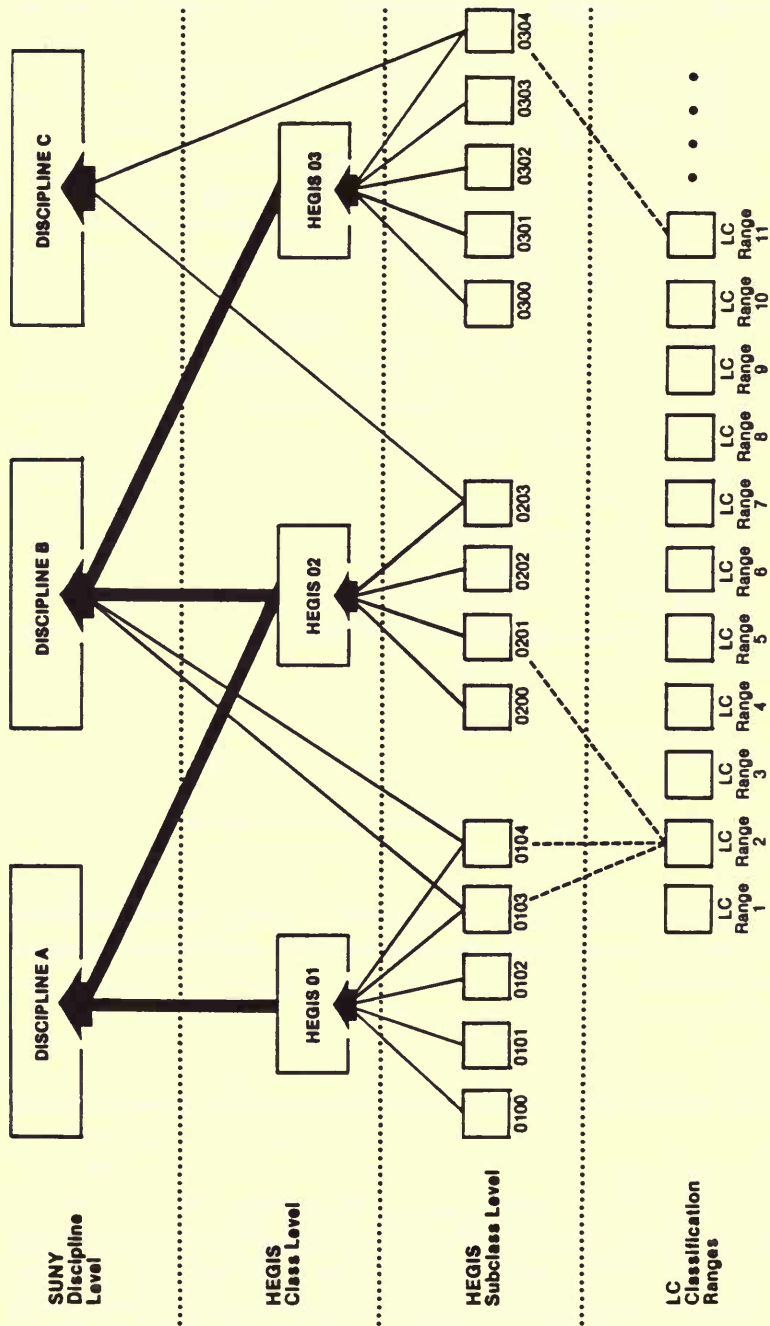


Fig. 3. Representation of the Mechanism by which LC Classifications are Assigned to a HEGIS Subclass and Higher Levels

DT 730-995	History, South Africa
G 2425-2739	Africa, Atlases
G 8200-8903	Africa, Maps
GA 75-76	Africa, Mathematical Geography
GA 286-286	Africa, Cartography
GA 1341-1673	Africa, History of Map Production
GB 330-378	Africa, Physical Geography
GB 439-440	Africa, Geomorphology
GV 135-143	Africa, Recreation
GV 1705-1713	Africa, Dancing
HA 1951-2275	Africa, Statistics
HC 501-591	Africa, Economic History
HD 1169	Africa, Land Tenure

Fig. 4. Sample of Entries from the African Studies HEGIS Class

enrollment data—specifically student credit hours per HEGIS class; (e) performing overlap studies within each discipline at each level using OCLC numbers; and (f) reporting by institution and level, listing titles in brief entry of required listed titles.

Steps which are to be added to the process to complete the programming of the model consist of: (1) incorporation of publishing data; (2) incorporation of use data; (3) development or acquisition of decision support system software and evolution of system to an online interactive state.

The Results

Two analytic processes—components of the total model—are now operational at the SUNY Central Computing Center. They are component analysis and overlap analysis. Component analysis is a process in which catalog records from OCLC tapes are passed against the HEGIS/LC tables

and allocated to subclasses, and the upward aggregates of class, discipline and institution. The analyses may also be correlated with the number of student credit hours taught in the discipline at the specific campus, based on CASA file data.

Overlap analysis occurs when the cooccurrence of the same OCLC number among the OCLC tapes of different campuses is used as a measure of uniqueness and commonality of holdings, and the grades between (e.g., held in three out of ten campuses). The overlap analyses are performed at the institution level, or, following a component analysis, at subclass, class or discipline level.

Results from these computer processes are designated as "obtained" results. The obtained results may themselves be subject to subsequent analysis, review and combination—as indeed they would be in a decision support system—to generate "derived" results. A start has been made on the process of producing derived results but by using the offline SPSS (Statistical Package for the Social Sciences) or manual analyses. Both obtained and derived results are reported in this paper. The derived results are sufficiently valuable to justify the target of achieving operational access to DSS software.

One major series of reports has been produced for the state university campuses, with data derived from those studies being the basis of all the results reported in this paper. In this series, the OCLC catalog tapes of eleven colleges of arts and sciences (four-year colleges), for the period April 1977 to December 1979, most with a small percentage of graduate (masters) programs were used. From this database, the study selected the latest use of monograph records which had a transaction code of *produce* and an imprint date of 1977 or 1978. This resulted in a base of 105,003 records for analysis. The attempt clearly was to gain an understanding of current acquisition decisions in the colleges. Other selections from the database could have as easily been made, ranging from the whole database to serials, updates and products. For our purpose, we chose the database we needed.

In the component analysis, the data were matched with student credit hour (SCH) data from the CASA file for 1978 (for which data are collected in the third week of the fall semester). We were exploring the academic demand at the campus for the 1978-79 academic year, matched with the acquisition of current (1977 or 1978) imprints which were received and cataloged by the library between April 1977 and December 1979. The data were analyzed first by the component analysis method, and then the overlap analysis method at all levels.

Component Analysis-Obtained Results

Figure 5 is representative of a typical page from a computer printout of the result of a component analysis. The four columns are respectively

the number of titles, the percent of titles allocated to each HEGIS/LC class for one campus from the database under review, the number of student credit hours and the percent of student credit hours for that campus.

The reader is cautioned that the first two, and the last two columns are added differently. The CASA/SCH data is a simple arithmetic sum. Either the students are enrolled in, say, a three-credit hour course, or not. However, because of the ground rule in the creation of the HEGIS/LC tables that it is possible for a class number to be assigned to more than one HEGIS class, it is perfectly possible and reasonable for a specific title to be allocated to more than one subclass in any one analysis. It is necessary for that multiple allocation to be removed at each step of the upward aggregation in order to avoid misleading and inflated results. Thus any multiple allocation of a specific record title will show as supporting the subclasses assigned, but will only contribute once to the class. A similar removal of duplication occurs at the upward aggregations from class to discipline, and discipline to institution.

This important point is illustrated in figure 5. The correct arithmetic sum for the number of titles in the class "Letters" is 4594, yet the reported number of titles is 4147. Given the consonance of the subclasses in the group, it is not surprising that such a multiple allocation can occur. Similarly, when the classes "Letters" and "Foreign Languages" are combined into a single discipline, the reported number is 5049, but the arithmetic sum ($4147 + 1511$) is 5658. This phenomenon also indicates an important practical consideration. If it is possible to identify for any one campus the subclasses and classes in which multiple allocation is taking place, the books which are bought are obviously lower-risk investment items than special areas of unique allocation—a nontrivial consideration in times of fiscal crisis. Finally, in column three the percent of titles is subject to the same rules of multiple allocation as the count of titles.

Component Analysis-Derived Results

One question raised by the component analysis is, simply: What is the percentage allocation between the SCH and the current acquisitions, and the ten-discipline HEGIS grouping? This allocation can be seen simply by charting the ten points for each discipline on a graph containing both acquisitions and the SCH. The results are demonstrated in figure 6. The result, which was startling, is that the graphs for ten of the eleven campuses were fundamentally the same as those shown by the three campuses in figure 6. The eleventh campus, SUNY at Purchase—the exception—is a new campus still busily building its basic collection. Thus the result is not surprising.

SUNY Grouping HEGIS Group/SubGroup	No. of Titles	% of Titles	No. of Student Credit Hours	% of Student Credit Hours	SUNY Discipline Totals HEGIS Class Totals
Letters and Foreign Languages	5,049	21,500	19,346.00	17,800	
11 Foreign Languages	1,511	6,430	3,130.00	2,800	
1101 Foreign Languages, Gen.	706	3.00	93	.08	
1102 French	172	.73	774	.69	
1103 German	94	.40	183	.16	
1104 Italian	116	.49	105	.09	
1105 Spanish	252	1.07	1,673	1.49	
1106 Russian	36	.15	36	.03	
1107 Chinese	10	.04	21	.01	
1108 Japanese	9	.03			
1109 Latin	4	.01			
1110 Greek, Classical	22	.09			
1111 Hebrew & Semitic	17	.07	78	.06	
1112 Arabic	5	.02			
1114 Scandinavian Languages	1	.00			
1115 Slavic Lang. Except Russian	9	.03	3		
1116 African Languages	11	.04	9		
1191 Hungarian	1	.00			
1193 Portuguese	44	.18			
1196 Persian	5	.02			
1197 Semitic Languages	17	.07			
1199 Foreign Languages, Other	13	.05	155	.13	
15 Letters	4,147	17,660	16,339.00	14,610	HEGIS Class Totals
1500 Letters Unclassified			57	.05	
1501 English General	190	.80			
1502 Literature, English	819	3.48	8,118	7.26	
1503 Comparative Literature	4	.01			
1504 Classics	81	.34			
1505 Linguistics	420	1.78			
1506 Speech, Debate, Forensic	19	.08	6,138	5.49	
1508 Teaching English as Forl	240	1.02	24	.02	
1509 Philology	456	1.94			
1510 Religious Studies	478	2.03	2,001	1.79	
1593 Technical Writing	10	.04			
1595 Children's Literature	219	.93			
1596 Literature of the Stage	654	2.78			
1597 American Literature	844	3.59			
1598 Mythology	174	.74			
1599 Letters, Other	5	.02			
1599 Letters, Other	1	.00			

Fig. 5. Sample Computer Printout of a Component Analysis

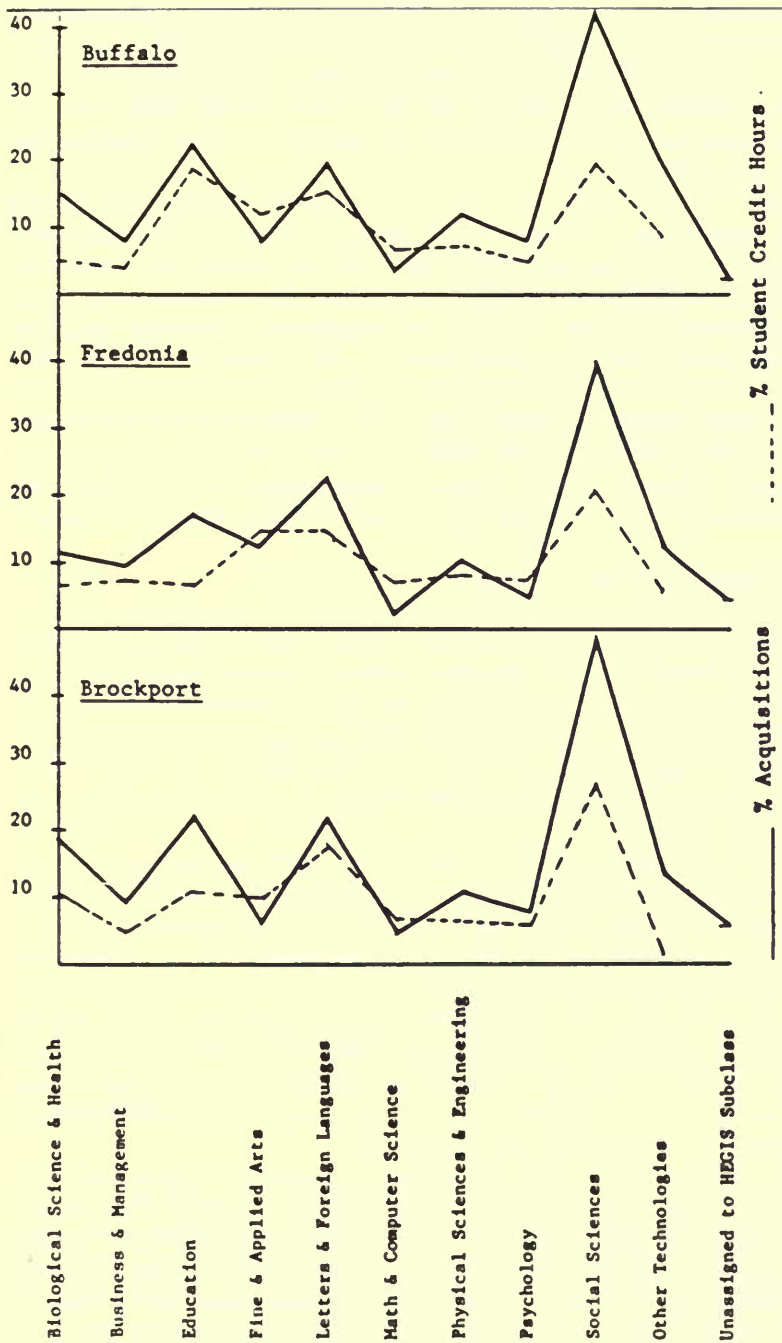


Fig. 6. SUNY Disciplines, Enrollment (Expressed as Percent of Student Credit Hours) and Percent of Total Acquisitions (having a 1977-78 Imprint Date) Assigned to the Discipline

In another test, scatter-grams were performed to examine the possible relationships between total credit hours and total titles acquired for each HEGIS class. In a scatter-gram, the coincidence of the values for each institution entered are displayed on the X and Y axis. The more the locations tend toward a straight line, the stronger are the relationships between the elements.

Table 1 shows areas for which very strong relationships existed between acquisitions and student credit hours. These eleven disciplines are now described as "immanent" program areas, proposing that there is an inherent relationship between the two elements. The third test undertaken with component analyses data was to examine the degree of multiple allocations among HEGIS classes, following the possibility of high *v.* low risk acquisition investment. The higher the index of multiple allocation of a title to more than one class, the higher the probability that the item will be used. An index of less than one would indicate a low correlation between the academic program and the acquisition program, diminishing as the index decreases. Table 2 shows the indexes for eleven SUNY campuses.

Overlap Studies—Obtained Results

The overlap studies (see table 3) examine the cooccurrence of titles among libraries based on the OCLC number at the subclass, class, discipline, and institution levels. The programs are designed for a maximum of ten institutions. This decision was frankly a programming compromise to obtain results quickly by avoiding the delay caused by the complexity of handling 100 institutions, as originally proposed. It has been found, however, that there is so little overlap beyond ten institutions that there may be little lost.

The results are displayed in a matrix in which the one column is the individual institutions identified by their OCLC codes with the total titles and copies in the last column. One column indicates the ten occurrences from unique (i.e., held by that institution uniquely), two (i.e., the institution plus one other), to ten (i.e., held by all institutions). The final column is the total of all titles. Each box in the matrix records the number and the percentage in each column. Table 3 reports the overlap for one campus and the total of all ten campuses among the subclass, class, discipline, and institution level. The holdings of one campus will not necessarily follow the pattern of the aggregate of all campuses.

The "total held by class" column is the sum of all copies held by a particular distribution—e.g., held in two or five libraries. The "actual titles" column is the number of titles which overlapped. This is best demonstrated in the ten-overlap column, where clearly the five titles held by ten libraries in the subclass will yield fifty copies (see table 3). Figure 7 describes graphically the overlap at the institution level found in this series of tests.

TABLE 1
MID-RANGE CLUSTERS AND THE PROGRAM AREAS
FOR WHICH STRONG RELATIONSHIPS WERE FOUND
BETWEEN ACQUISITIONS AND STUDENT CREDIT HOURS

<i>Program Area</i>	<i>r*</i>	<i>p</i>	<i>SUNY Colleges</i>		
SUNY—Biological Science & Health Professions	0.80 @	.004	Brockport Fredonia Oswego	Buffalo Geneseo Potsdam	Cortland New Paltz
Biological Science	0.97 @	.001	Buffalo Geneseo Plattsburgh	Cortland New Paltz Purchase	Fredonia Oswego
Health Professions	0.98 @	<.001	Brockport New Paltz	Fredonia Potsdam	Geneseo
SUNY—Fine & Applied Arts	0.86 @	.007	Brockport Geneseo Plattsburgh	Cortland New Paltz Purchase	Fredonia Oswego
Letters	0.76 @	.009	Buffalo Geneseo Plattsburgh	Cortland New Paltz Purchase	Fredonia Oswego
SUNY—Math & Computer Science	0.91 @	<.001	Buffalo New Paltz Plattsburgh	Cortland Oneonta Purchase	Fredonia Oswego
Mathematics	0.84 @	.009	Buffalo New Paltz Purchase	Cortland Oneonta	Fredonia Oswego
Physical Sciences	0.93 @	.001	Buffalo New Paltz	Cortland Oneonta	Fredonia Oswego
SUNY—Social Sciences	0.87 @	.003	Cortland Oneonta	Fredonia Oswego	New Paltz Purchase
Economics	0.88 @	.002	Buffalo Oneonta Purchase	Cortland Oswego	New Paltz Potsdam
Political Science	0.87 @	.006	Brockport New Paltz	Buffalo Oneonta	Cortland Plattsburgh

*r—Controlling for budget

It is one thing knowing the curve of the unique/commonality at the level described thus far; but it does not tell a library specifically how it relates to other libraries in the analysis. An extended report has been defined in which any one library can assess its overlap title-by-title with the

TABLE 2
INVESTMENT EFFICIENCY INDEXES FOR ELEVEN SUNY COLLEGES
OF ARTS AND SCIENCES—1977-78 IMPRINT DATE

<i>Institution</i>	<i>Index</i>
Brockport	1.86
Fredonia	1.67
Buffalo	1.81
Cortland	1.83
Geneseo	1.63
Oswego	1.74
Plattsburgh	1.77
Oneonta	1.70
New Paltz	1.75
Potsdam	1.51
Purchase	1.64
Mean	1.72

TABLE 3
OVERLAP STUDIES

<i>Institution</i>	<i>HEGIS Subclass-Afro American Studies 2211</i>		
	<i>Institution Code: XBM</i>	<i>Total Held by Class</i>	<i>Actual Titles</i>
Unique	23 17.42	68 9.48	68 33.66
2	21 15.90	72 10.04	36 17.82
3	16 12.12	54 7.53	18 8.91
4	10 7.57	48 6.69	12 5.94
5	14 10.60	85 11.85	17 8.41
6	11 8.33	72 10.04	12 5.94
7	12 9.09	91 12.69	13 6.43
8	11 8.33	96 13.38	12 5.94
9	9 6.81	81 11.29	9 4.45
10	5 3.78	50 6.97	5 2.47
Total	132 100.00	717 100.00	202 99.97

TABLE 3—Continued

<i>Institution</i>	<i>HEGIS Class—Social Science 22</i>		
	<i>Institution Code: XBM</i>	<i>Total Held by Class</i>	<i>Actual Titles</i>
Unique	3,214 36.41	9,091 22.29	9,091 51.95
2	1,615 18.29	6,234 15.28	3,117 17.81
3	1,106 12.52	5,178 12.69	1,726 9.86
4	806 9.13	4,620 11.33	1,155 6.60
5	671 7.60	4,065 9.97	813 4.64
6	470 5.32	3,354 8.22	559 3.19
7	376 4.25	3,087 7.57	441 2.52
8	296 3.35	2,576 6.31	322 1.84
9	181 2.05	1,647 4.03	183 1.04
10	92 1.04	920 2.25	92 .52
Total	8,827 100.00	40,772 100.00	17,499 99.97

TABLE 3—Continued

<i>Institution</i>	<i>SUNY/CASA Discipline—Social Sciences</i>		
	<i>Institution Code: XBM</i>	<i>Total Held by Class</i>	<i>Actual Titles</i>
Unique	4,228	11,262	11,262
	37.88	22.97	52.52
2	2,176	7,850	3,925
	19.49	16.01	18.30
3	1,356	6,222	2,074
	12.15	12.69	9.67
4	965	5,452	1,363
	8.64	11.12	6.35
5	798	4,825	965
	7.15	9.84	4.50
6	545	3,912	652
	4.88	7.97	3.04
7	439	3,598	514
	3.93	7.33	2.39
8	336	2,920	365
	3.01	5.95	1.70
9	213	1,953	217
	1.90	3.98	1.01
10	103	1,030	103
	.92	2.10	.48
Total	11,159	49,024	21,440
	100.00	100.00	99.60

TABLE 3—Continued

<i>Institution</i>	<i>Institution-Ten SUNY Colleges</i>		
	<i>Institution Code: XBM</i>	<i>Total Held by Class</i>	<i>Actual Titles</i>
Unique	8,241	25,474	25,474
	37.80	24.26	53.88
2	4,167	17,256	8,628
	19.11	16.43	18.25
3	2,642	13,281	4,427
	12.11	12.64	9.36
4	1,938	11,776	2,944
	8.88	11.21	6.22
5	1,521	10,035	2,007
	6.97	9.55	4.24
6	1,166	8,700	1,450
	5.34	8.28	3.06
7	884	7,224	1,032
	4.05	6.87	2.18
8	666	5,776	722
	3.05	5.50	1.52
9	405	3,771	419
	1.85	3.59	.88
10	171	1,710	171
	.78	1.62	.36
Total	21,801	105,003	47,274
	100.00	100.00	99.95

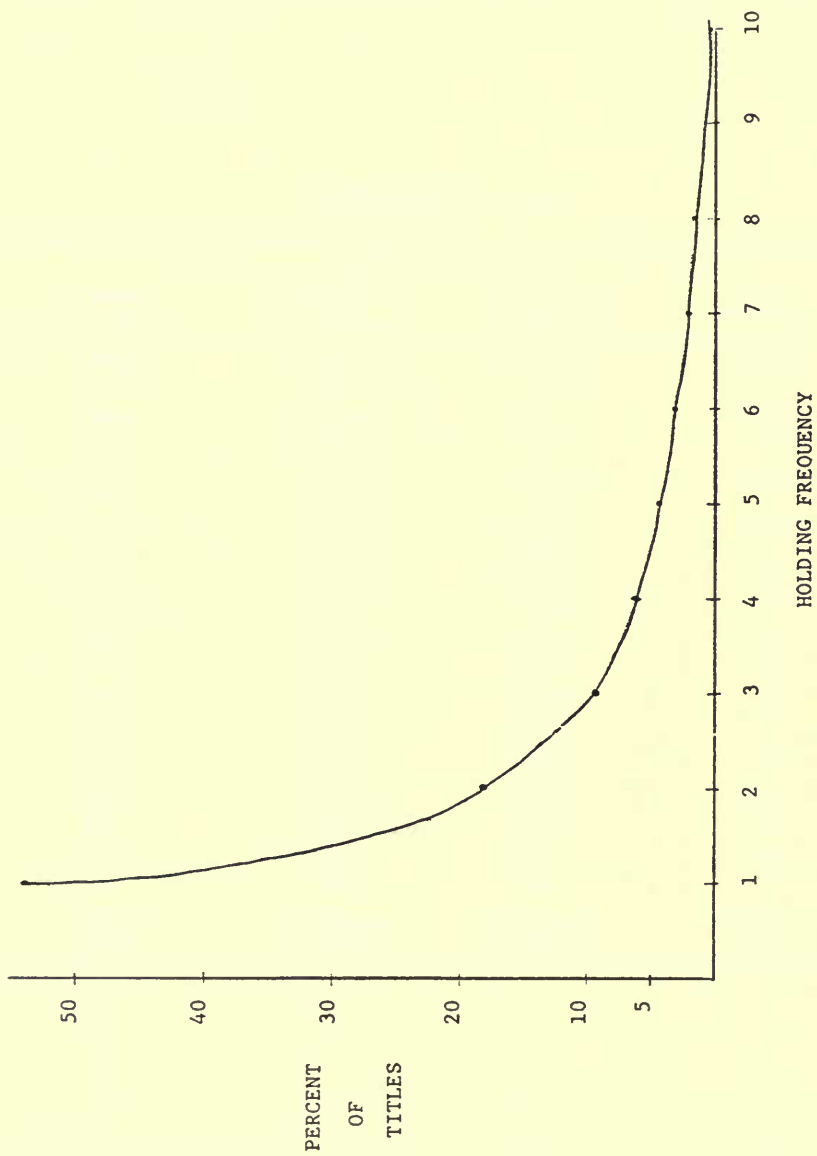


Fig. 7. Graph of Overlap of Titles at the Institution Level

other libraries. Figure 8 compares one SUNY library (Cortland) with the other nine in the analysis.

The last obtained report is a simple list of OCLC number of titles analyzed, with the OCLC holdings library code arranged alongside the number. This list is used as a crosscheck against the total results, but is also indicative of the ability of the system programs to produce bibliographies by library/libraries, by class, or by discipline. This facility has already been used to produce a bibliography of holdings in African and Afro-American studies for the chancellor's task force.

Overlap Studies-Derived Results

The overlap relationships can be described by five curves of highly unique, scarce, moderate, common, and ubiquitous distributions (see figs. 9 through 13). The distribution of highly unique material in figure 9 indicates a high proportion of unique material (usually 40 to 75 percent). Presumably this indicates a strong or specialized collection. The distribution of scarce material in figure 10 indicates a smaller proportion of unique items; nevertheless a large proportion of the collection is composed of material held by three or fewer libraries. Moderate distribution may be characterized by curves of several shapes (two possibilities are shown in figure 11). It indicates that the largest proportion of the collection is composed of materials held by from three to seven institutions. The distribution of common material may be characterized by several shapes (two possibilities are shown in figure 12). It indicates that approximately one-third of the materials are held by five to seven institutions and less than 10 percent are held by eight to ten institutions. The distribution of ubiquitous material may be characterized by curves of several shapes (two possibilities are shown in figure 13). It indicates that at least 10 percent of the material is held by eight to ten institutions.

One interesting area of study was the review of the uniqueness curves by the HEGIS classes. Table 4 reports the distribution among the SUNY libraries. Note that the lowest figure for uniqueness is psychology (47 percent), and that area studies (49 percent) is the only other figure below 50 percent. Note also that the number of titles at the bottom of column two, 47,274, is equal to the number of "actual titles" in the last section of the institution overlap study in table 3.

A review of the "scarce" titles by number and percent for each SUNY institution by HEGIS class is represented in table 5. These data are helpful in program and acquisitions review.

INSTITUTION YCM CORTLAND

SUBCLASS: 0305 African Studies

Total Held in Common With:

XBM = 80

XFM = 33

YBM = 60

YGM = 52

YPM = 44

YOM = 67

ZBM = 41

ZLM = 78

ZPM = 60

Total Unique = 15

Fig. 8. Volumes Cortland Library Holds in Common with Other SUNY Libraries

Conclusions

Work to Date

A model has been established to describe the library acquisition/retention process, and to support and inform that decision through the manipulation of machine-readable data derived as a byproduct of library, publishing and network operations. Where data are available, programming has been completed to perform disciplinary and overlap analyses on library holdings as recorded on OCLC tapes. Conversion tables from LC to HEGIS have been established.

A set of data from SUNY campuses have been analyzed through the programs and the results subjected to further review. These subsequent studies to achieve "derived" results were performed by SPSS and manual

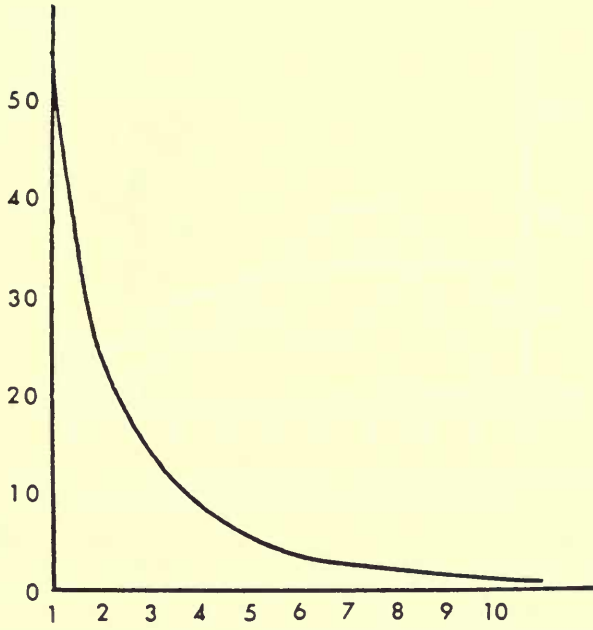


Fig. 9. Graph of Distribution of Highly Unique Material

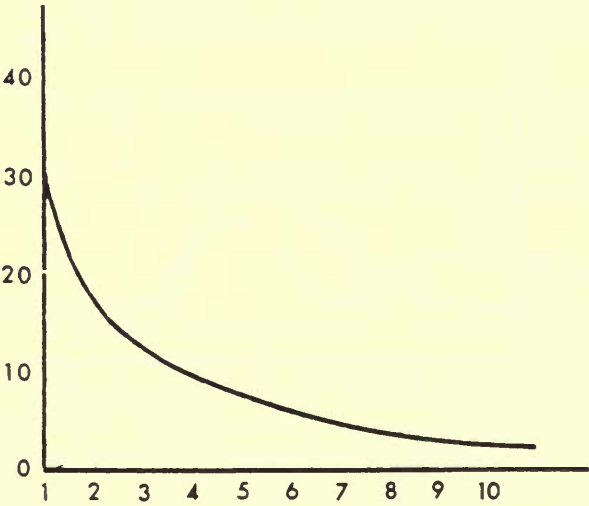


Fig. 10. Graph of Distribution of Scarce Material

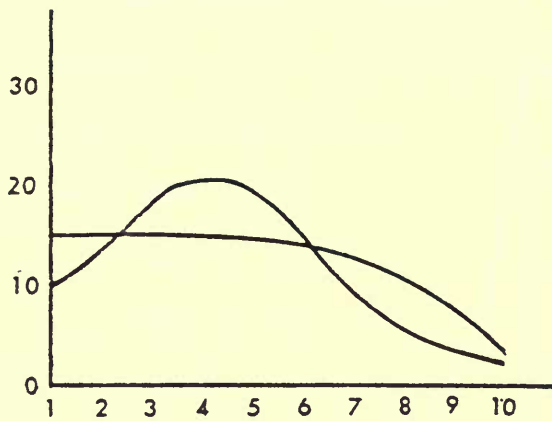


Fig. 11. Graph of Moderate Distribution

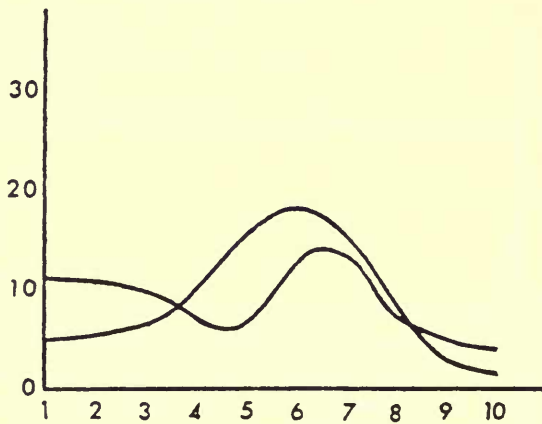


Fig. 12. Graph of Distribution of Common Material

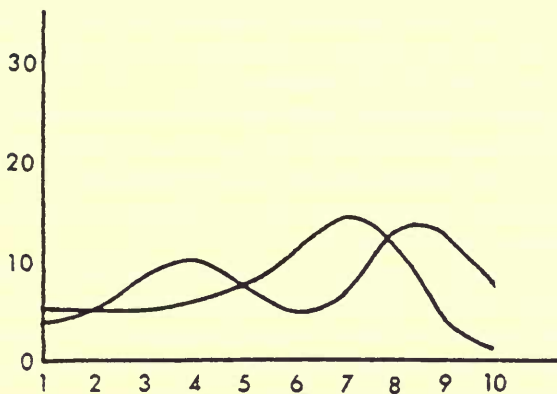


Fig. 13. Graph of Distribution of Ubiquitous Material

analysis, but they are amenable to DSS software when that becomes available.

Study of the results lead to some interesting conclusions. First, there is a very high degree of consonance in the *macrodecisions* made by campuses as they select materials. High correlations between discipline demands and acquisition decisions are observed within each campus. Some disciplines are described as being immanent, that is, they have an inherent relationship between program and acquisition patterns. We do not yet understand why these phenomena occur.

On the other hand, at the *microdecision* (i.e., the decision whether or not to acquire a specific title), remarkable diversity is shown even where programs are apparently similar among campuses. Only two disciplines which have a less than 50 percent uniqueness among ten campuses were found. These results point to four further conclusions:

1. It seems unlikely that an acquisition formula can be defined. It does seem probable that given the data available in this report, plus the incorporation of publishing and price data, a statement which describes the campus academic mission bibliographically can be produced and fine tuned over a period of time and in response to changing circumstance, and provide a firm justification for acquisition budgets.
2. The access system among campuses is essential. All must belong to the same bibliographic network, maintain their database, use the ILL message network, and the same document delivery system. The diversity among the campuses is the greatest bibliographic asset owned by the

university. Easy online access and rapid twenty-four hour delivery of documents is the best possible response to fiscal crises. This is, of course, true of any group of libraries which can display the same characteristics as the SUNY libraries.

3. The bibliographic diversity is a direct result of the subtle diversities of the academic programs at each campus, and a reflection of the book selection processes at each campus. The bibliographic description of the academic mission is a valuable contribution to the academic planning of each campus, and the university as a whole.
4. Although the results reported are based on current acquisitions, the university libraries at SUNY are engaged in retrospective conversion projects. Access to the whole of the bibliographic resource is the natural concomitant of the strategy proposed in number two above.

Planned Developments and Extensions

First, the studies reported in this study must be repeated at a later date to compare the results. The same criteria will be used in the repeat studies. Second, the analysis service is now being offered by SUNY/OCLC as a service to groups of libraries in the state and the country. A number of such analyses have been performed. Their results so far confirm the results reported in this paper. As more results are obtained, SUNY will accumulate them to build a broader picture and perhaps act as a clearinghouse for such studies. Third, Bowker has now announced the availability of its ABPR and BIP files for such projects as the SUNY project. They must be incorporated in the model. Fourth, the serials database in OCLC is growing rapidly. They must be incorporated in the study. Some preliminary studies are promising. Fifth, DSS software must be incorporated to extend the availability and utility of the derived data. Sixth, use data, primarily from OCLC it is anticipated, must be incorporated in the program. Seventh, the U.S. Office of Education has proposed a new and different set of HEGIS codes. Programs must be revised to accept these codes if they are to be accepted by the academic community. We do not have the cost, a time frame, or any sense of the ultimate improvement in the cost efficiency of this step, but if the data produced by the academic institution changes, we have no option. Eighth, ways must be found to incorporate the data into academic program planning within institutions. Ninth, the SUNY/OCLC Office of Library Services will seek to extend the utility of the analyses service and acceptance of the reports.

Prognosis

Progress on this project has been very slow. It has been difficult to attract support, increasingly so as the fiscal situation has deteriorated over

TABLE 4
 DISTRIBUTION BY HOLDING FREQUENCY OF
 1977-78 TITLES HELD BY TEN SUNY COLLEGES
 "IN TWO DIGIT HEGIS CLASSES"

<i>HEGIS Class</i>	<i>No. of Titles</i>	<i>Unique</i>	<i>Holding Frequency</i>			
			<i>2 or Less</i>	<i>3 or Less</i>	<i>5+</i>	<i>8-10</i>
22 Social Sciences	17,499	52%	70%	80%	14%	3.6%
39 Education	10,324	54%	74%	84%	11%	2.5%
15 Letters	8,885	52%	69%	78%	15%	3.3%
21 Public Affairs	7,172	54%	74%	84%	11%	2.7%
05 Business Management	4,316	57%	75%	84%	10%	2.3%
09 Engineering	4,021	62%	81%	90%	6%	0.8%
04 Biological Science	3,893	57%	75%	85%	9%	1.6%
24 Forestry	3,474	61%	79%	87%	8%	1.5%
10 Fine & Applied Arts	3,398	50%	69%	78%	15%	3.5%
11 Foreign Language	3,044	57%	72%	81%	14%	2.5%
20 Psychology	2,716	47%	64%	76%	16%	4.7%
49 Interdisciplinary	2,599	65%	80%	87%	8%	1.2%
13 Home Economics	2,502	52%	71%	80%	13%	4.0%
19 Physical Science	2,329	57%	76%	85%	9%	1.4%
12 Health Professions	1,980	63%	80%	89%	7%	1.5%
14 Law	1,862	75%	88%	92%	5%	1.4%
01 Agriculture	1,628	64%	81%	88%	6%	1.0%
02 Architecture	1,600	65%	80%	88%	8%	1.0%
06 Communications	1,329	50%	66%	77%	17%	3.3%
17 Mathematics	1,257	51%	77%	88%	4%	1.4%
03 Area Studies	1,249	49%	66%	76%	18%	3.2%
16 Library Science	1,116	56%	73%	85%	7%	1.0%
18 Military Science	799	52%	70%	78%	15%	2.7%
07 Computer Info. Sci.	563	62%	80%	91%	4%	1.4%
Mean (n = 24)		57%	75%	84%	10%	2.2%
Total Acquisitions	47,274*	54%	72%	81%	12%	2.8%

*Note: Because this table is the result of a component analysis, in which one book may be assigned to more than one discipline, 47,274 is the number of titles assigned to disciplines and is not the sum of all assignments.

TABLE 5
 PERCENT OF SCARCE TITLES AND NUMBER OF TITLES PER HEGIS CLASS BY INSTITUTION
 1977-78 IMPRINT DATES*

	Brockport	Fredonia	Buffalo	Cortland	Geneseo	Plattsburgh	Oswego	Oneonta	New Paltz	Purchase
22 Social Science	67%	52%	53%	50%	47%	34%	45%	52%	43%	30%
	8827	2637	3868	4174	4128	2438	3855	4754	3881	2210
08 Education	74%	44%	58%	64%	47%	38%	51%	57%	45%	33%
	5150	1303	2404	2693	2187	1175	2057	2613	1902	574
15 Letters	63%	38%	55%	34%	41%	28%	57%	56%	42%	30%
	4127	1409	1860	1484	2060	1255	1976	3216	2383	1329
21 Public Affairs	75%	36%	58%	65%	44%	36%	49%	56%	41%	29%
	4012	783	1713	2239	1359	759	1244	1670	1191	527
05 Business/Management	74%	55%	59%	57%	58%	43%	57%	57%	46%	29%
	2090	716	883	769	959	534	913	1096	684	312
20 Psychology	69%	26%	48%	38%	36%	30%	40%	43%	37%	22%
	1787	383	852	600	650	416	599	796	628	223
49 Inter. Discipline	77%	57%	62%	64%	59%	51%	62%	64%	51%	40%
	1099	291	489	476	458	265	562	617	349	223
13 Home Economics	61%	28%	61%	50%	43%	34%	48%	60%	40%	28%
	1063	282	824	668	581	302	554	854	501	215
19 Physical Science	71%	64%	59%	50%	69%	65%	61%	61%	50%	38%
	894	451	474	274	722	248	582	523	310	174
12 Health Professions	82%	52%	61%	73%	55%	56%	61%	66%	45%	46%
	1067	177	318	573	328	197	271	366	241	101
09 Engineering	77%	66%	73%	69%	70%	58%	68%	71%	61%	48%
	1401	540	901	565	889	349	880	918	533	244
04 Biological Science	77%	51%	63%	63%	55%	49%	59%	58%	49%	44%
	1401	438	762	816	671	549	799	783	612	322

TABLE 5—Continued

	Brockport	Fredonia	Buffalo	Cortland	Geneseo	Plattsburgh	Oswego	Onondaga	New Paltz	Purchase
24	81%	54%	64%	63%	61%	54%	60%	61%	57%	47%
Forestry	1779	355	655	581	592	395	620	734	531	290
10	57%	43%	51%	33%	49%	21%	46%	45%	52%	54%
Fine/Applied Arts	1241	555	797	487	813	435	519	987	1203	1111
11	72%	44%	51%	44%	43%	29%	52%	52%	47%	35%
Foreign Language	1506	488	527	469	502	413	657	904	885	447
14	73%	44%	74%	90%	80%	24%	62%	67%	43%	30%
Law	392	108	357	810	427	92	238	295	143	76
01	76%	44%	72%	68%	64%	51%	65%	74%	64%	41%
Agriculture	634	142	354	307	235	178	316	431	231	111
02	82%	44%	64%	61%	56%	38%	61%	62%	56%	57%
Architecture	808	106	333	207	245	104	271	369	273	214
06	60%	35%	50%	37%	43%	27%	52%	49%	48%	36%
Communication	620	198	362	230	354	198	291	430	425	146
17	87%	59%	58%	31%	68%	52%	70%	64%	64%	25%
Mathematics	831	121	234	94	366	83	322	187	166	32

•Note: The first set of numbers in each case is the percent of titles in the HEGIS class which are scarce (held by three or fewer institutions). The second set is the total number of titles held in the HEGIS class. Thus 67%/8827 indicates that of 8827 titles, 67% are scarce.

the last few years. At the same time, although the service is available, few libraries have taken advantage of it. There is also difficulty in obtaining acceptance and integration of the results into the academic and bibliographic decision-making process. It just takes time and patience.

It is, however, inevitable that, because of their fitness and because of the increased pressures and complexities of decision-making, automated library modeling systems, supported by the analysis of library and other operational data, will gain slow, reluctant acceptance by administrators, budget officials and librarians.

REFERENCES

1. "The Shorter Catechism of Stafford Beer." *Datamation* 28(Feb. 1982):148.
2. Spaulding, F.H., and Stanton, R.O. "Computer-aided Selection in a Library Network." *JASIS* 27(Sept./Oct. 1976):269-80.
3. Keen, Peter G.W. "Information Systems and Organizational Change." *Communications of the Association of Computing Machinery* 24(Jan. 1981):24-33.
4. King, William R. "Using and Evaluating Administrative Decision Support Systems." *New Directions for Higher Education* 35(Sept. 1981):64.
5. Hopkins, David S.P., and Massy, William F. *Planning Models for Colleges and Universities*. Stanford, Calif.: Stanford University Press, 1981, p. 4.
6. King, "Using and Evaluating."
7. Clapp, Verner W., and Jordan, Robert T. "Quantitative Criteria for Adequacy of Academic Library Collections." *College & Research Libraries* 26(Sept. 1965):371-80.
8. New York State Education Department. *Guidelines for Assessing the Adequacy of Academic Libraries in New York State*. Albany, N.Y.: 1976.
9. Evans, Glyn T., et al. *Development of a Responsive Library Acquisitions Formula: Final Report*. Albany, N.Y.: SUNY, Central Administration, 1978.
10. Beazley, Richard M. *Library Statistics of Colleges and Universities: 1979 Institutional Data*. Washington, D.C.: National Center for Educational Statistics, 1981.
11. Beilby, Mary H., and Evans, Glyn T. "An Information System for Collection Development in SUNY: A Progress Report." *Collection Management* 2(Fall 1978):217-28.
12. Huff, Robert A., and Chandler, Marjorie O. *A Taxonomy of Instructional Programs in Higher Education*. Washington, D.C.: National Center for Educational Statistics, 1970.
13. Office of Institutional Research and Analytic Study. *Statistical Abstracts* (Official report produced by the Course and Section Analysis system containing information on instructional workload, faculty, I&DR). Albany, N.Y.: SUNY, OIRAS, 1982.