A Critical Examination of the Assessment Analysis Capabilities of OCLC ACAS

by Lucy E. Lyons

Available online 28 October 2005

Over 500 libraries have employed OCLC's iCAS and its successor Automated Collection Assessment and Analysis Services (ACAS) as bibliometric tools to evaluate monograph collections. This examination of ACAS reveals both its methodological limitations and its feasibility as an indicator of collecting patterns. The results can be used to maximize the assessment capabilities of ACAS.

Lucy E. Lyons is Bibliographer for Collection Assessment and Planning, Northwestern University Library, Evanston, IL 60208-2300, USA <l-lyons@northwestern.edu>.

Introduction

This study, based on the experiences of Northwestern University Library's participation in the Illinois Statewide Assessment Project (ISAP), explores the feasibility of using Automated Collection Assessment and Analysis Services (ACAS) to analyze the monograph collections of academic libraries. It is not a report on the results of a particular study, but rather its purpose is to supply a descriptive and critical review of a collection assessment tool which, according to OCLC, has been purchased by over 500 libraries.

The short history of technological developments in evaluation tools parallels a long history of cooperative statewide collection assessments by college and university libraries in the State of Illinois. In 1986–1987, Northwestern University Library participated in the Illinois Collection Assessment Matrix (ICAM). This study utilized LCS shelflist tapes. The results were sent to each participating library as printouts.[1] Ten years later, in conjunction with other academic institutions in Illinois, the Library purchased the OCLC/Amigos Collection Analysis CD (CACD). There were several advantages to CACD over ICAM. Most importantly, it was computerized. The disadvantages included the fact that it was difficult to share the single disk between libraries and the program failed whenever changes to default parameters were attempted. These problems were not insurmountable, but some fixes were costly. In any case, in 1998 OCLC announced that CACD would be discontinued because "the application is older and increasingly difficult to maintain."[2]

In early 1999, the Western Library Network (WLN) merged with OCLC and developed the Automated Collection Assessment and Analysis Services (ACAS) with a new and searchable CD-ROM product called the Interactive Collection Analysis System (iCAS). The Illinois Statewide Assessment Project began, in 2000, to encourage academic libraries to join a study utilizing this evaluation tool.[3] With financial assistance from each participating library as well as the Illinois Cooperative Collection Management Program (ICCMP), eighty libraries joined together and agreed to submit their WorldCat bibliographic records to the project. The final iCAS statewide database contained information on monographs published up to early 2001 and held by these libraries. Before these data were systematically analyzed, iCAS was re-configured as a Web product, now simply referred to as OCLC ACAS. Northwest-

ern University Library, along with eighty-two others, engaged in this follow-up study. In just over fifteen years, the available tools for comparative collection assessment evolved from manually entered results of shelflist tapes distributed in paper; to a disk loaded onto a DOS platform; to searchable CD-ROMs residing in a Windows setting; to the current Web-based environment.

Background

There are numerous reasons to execute collection analyses in academic libraries. Grant applications submitted by academic programs sometimes require library statistics to prove that the university has the infrastructure and resources necessary to support the inquiry for which the proposal seeks additional assistance. In large institutions, academic departments are required to undergo periodic internal and external program reviews for which library statistics are included. And, of course, libraries themselves oftentimes undergo the same process. Collection assessments are also useful for forming a foundation for collection development planning; coping with budget shortfalls; fulfilling accreditation reportage; spotlighting cooperative collection sharing opportunities; justifying budget requests; and for accountability.

Most literature on collection evaluation concentrates not on the tools of analysis per se but on assessments of particular collections, from books, journals, and serials to electronic reserve, networked resources, and subject portals.[4] In contrast, with the exception of reports in regards to OCLC/Amigos CACD,[5] a thorough search of the library literature produced a surprising silence on the subject of OCLC assessment tools. While there is no report on the ACAS tool under question, there is one mention to its predecessor, iCAS. It did not explore iCAS as an evaluation tool or in terms of assessing library collections; rather, the software was applied to a random sample of OCLC records to test the expedience of employing WorldCat as a cooperative collection development resource.[6]

In consideration of the cost of the ACAS products as well as the cost in staff time to make use of them, one might expect to find at least a minimal amount of critique in the literature. Although many libraries have purchased the ACAS product herein described, there are several reasons that may account for the lack of literature as well as serve as a warning: the amount of data generated is immense; it is cumbersome to manipulate; and much time is required to procure the benefits.

With these facts in mind, a search was undertaken to find out if there have been any preliminary observations posted on the Web. Two consortia within the United Kingdom have reported their experiences in the use of the predecessor iCAS product. The most extensive is the final report of the "CURL/RSLP Collection Mapping Project based on OCLC/LACEY iCAS Software." Six institutions participated in this project to explore if iCAS could be fitted into a UK environment that uses classification systems outside of the standard U.S. schemes. It was found that, with adjustments from OCLC, iCAS was able to map up to 84 percent of non-LC and Dewey records. A second conclusion states that "the iCAS software can provide a fair, albeit incomplete and relatively blunt, analysis of collections described in online catalogues in terms of Conspectus categories."[7] The report further noted needs for refinement including (1) an examination of "No Call Number Present" records, (2) improvement in concert with OCLC on the Conspectus system, and (3) a test of the Web-based ACAS product (critiqued herein) in hopes of overcoming the problems of iCAS overlap and uniqueness title-level studies (which included "No Call Number" titles in the analyses).

The second UK report is from the "Mapping Medicine" pilot project involving eight biomedical libraries in the London area. The latest "progress report" available notes, in similarity to the CURL/RSLP study, that match rates with LC subject headings were improved following collaborative work with OCLC after "concerns were raised over the validity of the initial ACAS results." At the time of the report, successfully matched records had been raised in rate from 49 to 63 percent.[8]

Finally, the Orbis Collection Development Committee (CDC) of the University of Oregon reports in a preliminary investigation that despite "the useful nature of iCAS data ... the OCLC project is not without its drawbacks. Using the product effectively and interpreting the findings would require some training." It also noted "especially when the CDC factored in the long-term usefulness of the collected data." This concern is explained as such: "Because this compilation only includes records to date, within three to five years, the data is no longer current and the UO would have to expend another sizeable amount of money in order to update its information."[9]

The following section supplies a descriptive and critical overview of the ACAS bibliometric tool. It is succeeded by five segments that examine its main analytic features: size, age, and growth of collections, and title overlap and uniqueness. These sections explore the mechanics and logic of the product and, in the form of diagnostic questions and examples, suggest methodologies for maximizing ACAS's assessment capabilities. The tables demonstrate these aims; as their purpose is not the reportage of results, the statistics within the tables are generally hypothetical. The Conclusion contains the essence of the research results.

ACAS's Assessment Abilities

Although numerous analytic options are available through ACAS, the richness of analysis will depend on the wealth of the libraries involved. It supplies WLN Conspectus records in Dewey and LC for twenty-four broad subject "divisions" (e.g., Anthropology) as well as approximately 500 subject "categories" (e.g., Ethnology and Ethnography) within the divisions, and 4000 "subjects" (e.g., Culture and Cultural Processes) within the categories. Title-level, audience-level (adult versus juvenile readership levels), and language analyses are also available. Eighty libraries in Illinois joined together in purchasing the product, but economics ruled out title and audience-level analyses.

"The ease of current Web access to ACAS is a welcome improvement."

Like its successor, it was not difficult to create tables on iCAS, but the process of starting the CD-ROMS—all of the program's data and interpretive software—was very slow. The ease of current Web access to ACAS is a welcome improve-

ment. The landing page allows the user to choose either an analysis in Dewey or LC. It also provides demonstration projects for review. Once a classification system is selected, a "title analysis" or date-range table containing the holdings information of all participants appears. This is an important base table. Within this table resides the data from which three of the five major analytic capabilities of ACAS spring—namely, size, age, and growth of collections. Manipulation of the data from this table also provides the added-value analyses not readily available from any ACAS program features—e.g., descriptive statistics such as comparisons of relative strength within categories vis-à-vis distributions within the divisions of selected libraries. Of the two remaining major ACAS capabilities—title overlap and title uniqueness—each spring from tables with sets of frequency distributions that are different from each other and different from the date-range table.

In the initial "title analysis" table, the data of each library forms the rows. Date-of-publication ranges form the columns. In the case of the Illinois Statewide Assessment Project, there are eighty-six libraries listed (including four all associated with Northwestern University) and a massive, overwhelming amount of data. The "limit" feature allows the user to manage and select specific variables. It is possible to limit: the division (e.g., to select Sociology only); publication dates (e.g., 18th century); languages; and libraries.

The "settings" tab provides an opportunity to choose a type of analysis (the column data)—age or growth of the collections, a study of overlap, or comparisons of title uniqueness. It also allows the user to control variables such as division, category, subject, language, and library (the row data). A very important and useful feature permits the user to arrange the sequence of these row variables. Both the "limit" and "settings" features are very well designed, easy-to-use, and simple to execute.

Fig. 1 is a snapshot of an LC analysis produced by limiting the library to Northwestern University, the Division to Sociology, the date ranges to the 1930s and 1940s, and the languages to German and English. The ordering sequence was set to display the library first, then the division, languages, categories, and lastly the subjects. The latter two can be fully expanded to reveal the number of titles held by the Library in English and German, published in the 1930s and 1940s within subject areas of Sociology. The figure is a partial list of categories and counts of titles in German.

Fig. 2 provides a partial look at an analysis of unique titles limited to four libraries. This table demonstrates two notable facts alluded to earlier. First, the amount of data is massive. The table is but a glimpse of the true result. Online, the "shared by" columns stretch out to "shared by 73"; hence, including the total count of unique titles, there are seventy-four columns in all. Secondly, although the total of unique titles is supplied, no

Figure 1
Results Using the Limit and Settings Features

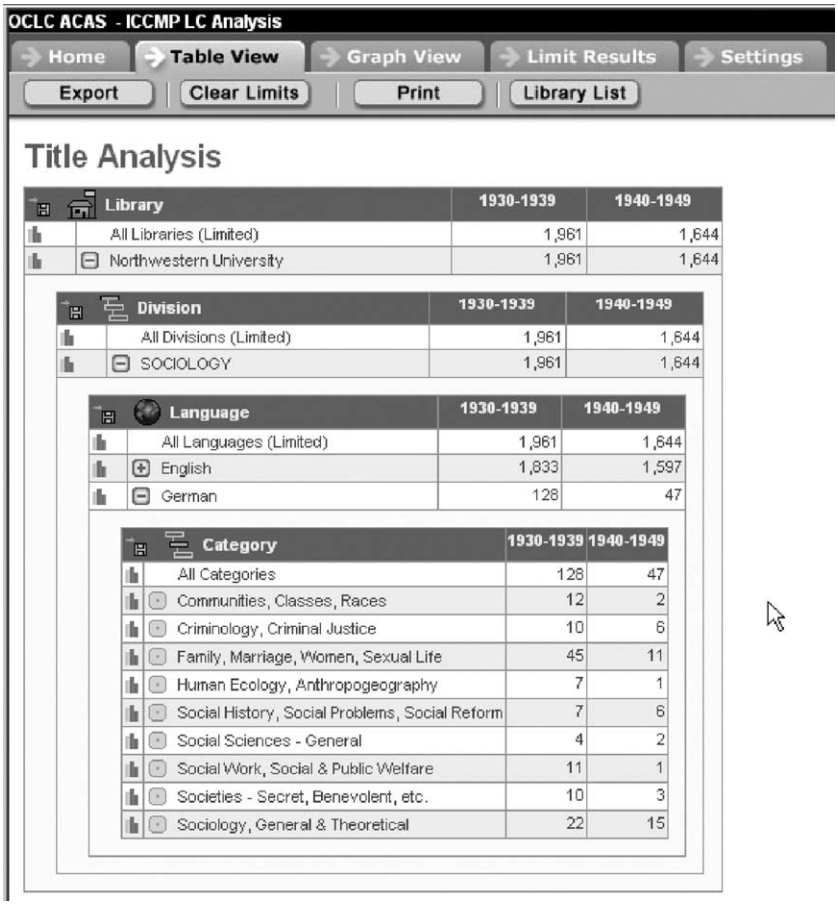
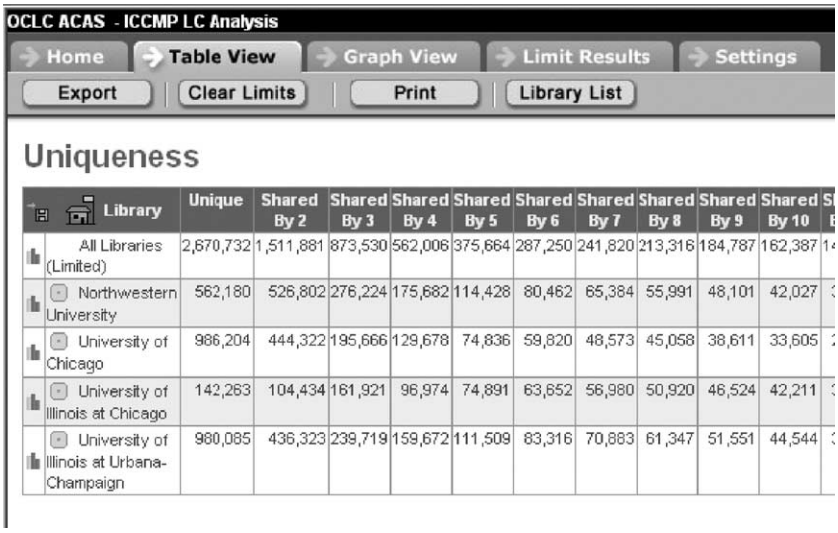


Figure 2
Partial Image of Analysis of Unique Titles Limited to Four Libraries



Library	Unique	Shared By 2	Shared By 3	Shared By 4	Shared By 5	Shared By 6	Shared By 7	Shared By 8	Shared By 9	Shared By 10
All Libraries (Limited)	2,670,732	1,511,881	873,530	562,006	375,664	287,250	241,820	213,316	184,787	162,387
Northwestern University	562,180	526,802	276,224	175,682	114,428	80,462	65,384	55,991	48,101	42,027
University of Chicago	986,204	444,322	195,666	129,678	74,836	59,820	48,573	45,058	38,611	33,605
University of Illinois at Chicago	142,263	104,434	161,921	96,974	74,891	63,652	56,980	50,920	46,524	42,211
University of Illinois at Urbana-Champaign	980,085	436,323	239,719	159,672	111,509	83,316	70,883	61,347	51,551	44,544

column provides overall total holdings of each library. While it is interesting to see that the University of Chicago is listed as having the greatest absolute number of unique titles, we cannot tell from this table if the percentage of unique titles to its total holdings is greater than those of the other institutions. With these caveats in mind, it is fortunate that ACAS provides an export feature.

The export function, like ''limit'' and ''settings,'' is flexible. The user may maintain or exclude saved limits, choose a ''conspectus dimension'' (division, category, subject), a ''descriptive dimension'' (language or library), and an analysis (date range, overlap, or uniqueness). The resultant table will export to Excel and save as a comma delimited file. In Excel, one may merge documents, sort, add or hide columns or rows, insert formulas, and create charts and graphs. It is also possible to reconfigure the subject divisions. At Northwestern University, for example, the Department of Economics and the School of Management are completely separate entities. To reflect this in an analysis of the Library's collection, the categories and subjects of the ''Business and Economics'' division were exported, sorted, and redesigned into two separate divisions—one called ''Business'' and another ''Economics.''

To mine the true richness of the data, the best excavation tool must be employed. Many of the following types of analyses are possible only if the data are exported and further manipulated using Excel application software.

Analysis 1: Size of Collections

The ACAS tool does not provide access to absolute collection size at most levels of analysis. At best, users can report the distribution of titles per library at the institution level; absolute collection size is not possible to gage at the division, category, or subject levels. At these levels, titles were mapped by call number and the OCLC matching algorithm program rejects records if any part of the string fails to match. All libraries have titles that do not have ''valid'' call numbers, especially large, academic libraries with older collections. While analysis on the total holdings level is unaffected, all other analyses levels are affected by the dropout of titles that have invalid call numbers. (Language is the one exception as it is pulled by the MARC tag for language rather than by call number.) Further, ''No Call Number Present'' counts will skew the results of overlap and uniqueness runs. As it cannot be assumed that such records did not match because they represent unique titles, it is very important to exclude these counts from overlap and uniqueness studies. There are many reasons for discrepancies and for rejections, but the end result is that approximately 20 percent of the records of the University of Chicago and 21 percent from Northwestern University cannot be analyzed beyond the absolute collection size for each institution.

''The ACAS tool does not provide access to absolute collection size at most levels of analysis.''

Concern over the effect of missing records led Northwestern University Library to compare ACAS results to data from the North American Title Count (NATC). In the subject area of ''Education,'' for example, the NATC provided a count of 74,197 titles while the ACAS count was 70,822. Confidence in the data was stabilized by the direction and proportion of the results. According to the NATC, ''Education'' titles are 4.01 percent of all titles in the entire collection; the proportion according to ACAS is 4.07 percent. Furthermore, the most prominent subfield or category in ''Education'' was found by both NATC and ACAS to be ''Higher Education'' at 38–39 percent, respectively. Not all subject areas in the two databases were so clearly comparable: confidence in the results was not great enough to use ACAS as an analysis tool for some subject areas including Geography and Music. (The reason for the

Table 1
Strength of Chemistry Collection Relative to Other LC Subject Divisions

Rank	Divisions	No. of Titles	Percentage of Total Holdings
1	Language, Linguistics, and Literature	418,968	24.10
2	History and Auxiliary Sciences	265,634	15.28
3	Business and Economics	201,064	11.57
4	Philosophy and Religion	93,531	5.38
5	Sociology	80,900	4.65
6	Engineering and Technology	76,549	4.40
7	Art and Architecture	75,142	4.32
8	Political Science	70,964	4.08
9	Education	70,822	4.07
10	Library Science, Generalities and Reference	53,355	3.07
11	Geography and Earth Science	52,619	3.03
12	Law	49,921	2.87
13	Physical Sciences	31,934	1.84
14	Medicine	31,112	1.79
15	Biological Sciences	29,761	1.71
16	Music	28,527	1.64
17	Mathematics	22,301	1.28
18	Psychology	17,199	0.99
19	Chemistry	16,853	0.97
20	Agriculture	15,292	0.88
21	Anthropology	14,500	0.83
22	Performing Arts	9900	0.57
23	Computer Science	6979	0.40
24	Physical Education and Recreation	4371	0.25
	Total	1,738,198	100.00

discrepancy in the case of the former is unknown; in the case of the latter, the problem may have been associated with musical scores.)

Further examination of the data showed that the dropout rate from "No Call Number" records did not skew all of the data of the Library beyond the point that it can be serviceable (albeit in a limited way). The Library has two major collections of distinction: Africana and Transportation. Throughout all analyses of the data, the uniqueness and subject strengths within these collections remained visible. In the case of Northwestern's collection, therefore, it was concluded that ACAS may be said to be useful as an indicator or as a sketch rather than as a complete portrait of the collection.[10]

With size as an indicator of collection strength, Table 1 is a visual analysis of the strength of the "Chemistry" collection of

Table 2
Ranking of Subfield Strengths within Chemistry Division

Categories	No. of Titles in Category	Percentage of Titles in Division
Physical and Theoretical	4195	24.89
Organic	3117	18.50
Analytical	3016	17.90
Chemistry, General	2241	13.30
Inorganic	2221	13.17
Crystallography	2063	12.24

a single library. The table shows that Chemistry is not a major focus of collecting relative to the other LC subject divisions. Calculate the mean number of titles (72,424) within the twenty-four divisions and Chemistry (with 16,853) proves to be weaker than the average collection at this library.

Table 2 illustrates ACAS's ability to look inside the Chemistry collection to see if there is any particular strength within. As the figures show, "Physical and Theoretical" chemistry has been a selection priority.

Of the total holdings in Table 1, less than 1 percent of the books are cataloged in Chemistry. Does Chemistry typically make up less than 1 percent of like-libraries' total holdings? In Table 2, almost one-fourth of the Chemistry collection contains books on "Physical and Theoretical" chemistry. Do other libraries have equivalent collection levels in this subject category? Fig. 3 is an example of ACAS's ability to answer these types of questions.

The bar graph indicates that the proportion of Chemistry holdings composed of titles devoted to "Physical and Theoretical" at Library 1 is about twice as high as all the others. Library 1 appears to have a specialization in this subfield compared with the other three libraries that have equivalent collection levels.

While it is possible to repeat many of these queries at the subject level, the analyst must decide when the title counts become too low to be relevant. Although several university libraries involved in the Statewide Assessment have substantial holdings, the frequency distributions at the subject level (with 4000 subjects) were often too low to be relied upon for making credible interpretive statements.

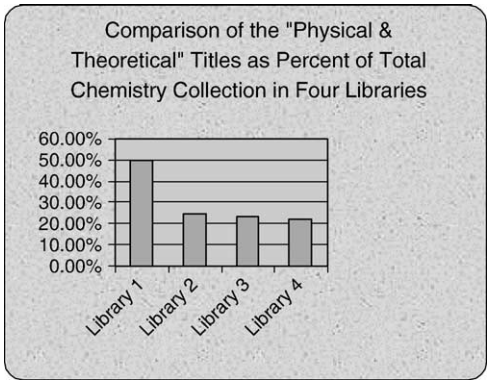
In short, depending on the library's analysis of "No Call Number Present," ACAS statistics on size might be used as indicators of collection strength and adequacy to answer questions such as follows: Where do the library's total holdings rank among peers? What are the strengths of the collection in terms of the twenty-four broad divisions? Are there clear collection strengths within each division (does the collection has a narrow focus)? Or, are there some divisions wherein the distribution of the collection is equally proportioned across categories (a broad focus)? How do these strengths compare with the same divisions within selected peer institutions? When the divisions are grouped into disciplines, in which is the collection weakest—Humanities, Sciences, or Social Sciences? As a percentage of the total holdings, what is the top twenty-five of the 500 categories in the library? What is the top 100 of the 4000 subjects? Are the bottom fifty category collections too small and inadequate for the libraries' needs? Compare collection priorities with selected peer libraries by sorting the data by category and ranking it within category by highest to lowest percent of the division. If "African Languages" is the strongest subfield of the library's "Languages and Literature" division, how significant does it remain when compared to like-holdings in peer libraries? Among small collections that are low selection priorities in your library, in which libraries do these appear as high priorities?

Analysis 2: Age of Collections

Publication date is used to measure the age of materials. Date information is gleaned from the MARC008 field, or, if not found in the fixed field, from the 269$c subfield. An analysis of date ranges within Northwestern University Library's holdings found that non-19th and non-20th century titles were dropped at a significantly higher rate in division and category analyses than more recent titles. That is, these older materials had the highest rates of "No Call Number Present." Although there were 101,095 titles listed as published prior to 1800, under subject analysis 68 percent were dropped: only 32,349 or 32 percent had "valid" call numbers. This is an exercise other libraries may want to repeat prior to analyzing older titles.

Frequency distribution tables revealed that in the case of the Statewide Assessment, at least 90 percent of each institution's collection is composed of monographs published in the 20th century. This figure was found by exporting the ACAS "date ranges" table, grouping all dates by century, and calculating the percentage of total holdings falling within each century for each institution. Given the fact that the vast majority of all library holdings are 20th century, ACAS's date ranges were found to be in general suitable and sensible. The ranges are grouped as such: all titles published prior to the year 1500 are grouped together; all 16th century combined together; all 17th century together; all 18th century together; 1800–1899 grouped per decade; 1900–1980 per decade; and 1981–2003 listed per year. Some large libraries might prefer all of the figures of the

Figure 3
Peer Comparison of Specialization in Chemistry Collections



20th century to be listed per year, leaving mergers to the user. In the case of the Statewide project, the last year (2003) included only partial holdings; hence, 2003 data had to be excluded when full year comparative analyses were performed.

"Frequency distribution tables revealed that in the case of the Statewide Assessment, at least 90 percent of each institution's collection is composed of monographs published in the 20th century."

These date spreads allow for many comparisons including: per century; per half centuries within the last two; per decade comparisons within the last two centuries; and 20th century holdings versus all others.

Because the proportion of titles published in the 20th century is so high, interpreting data on the older materials is not straightforward. With relatively few numbers spread throughout the centuries, it may be difficult to narrow down the importance of any particular collection. One method is to group date ranges together—e.g., all items published prior to the 19th century. Run an analysis on those selected dates by division. Retrieve the total holdings of the division (all date ranges) and calculate the percentage of pre-1800 publications within each division. This analysis showed that within Northwestern University's "Philosophy and Religion" division about 10 percent of the monographs were published before the 19th century.

One step further, taking the analysis down to the category level, and one is inundated with irrelevant results; nonetheless, an analysis which displays categories by division reveals what may otherwise be hidden. For this analysis, it is necessary to create the following table: number of titles in the category published prior to the 19th century; total holdings in the category—all dates; and percentage of pre-19th century holdings in each category. Further analysis of the Northwestern University Library's "Philosophy and Religion" division revealed that while overall almost 10 percent are pre-19th century monographs, 40 percent (2666 titles) of the "Protestantism" collection (6644 total titles) within this division is composed of this older material.

It is the norm with the ACAS program to face a cumbersome amount of data. Continuing with the example above, 32,349 pre-19th century titles spread over 500 categories produced only forty-six categories in which the total holdings were made up of 1.5 percent or more of older titles. In only fourteen categories do pre-19th century titles compose at least 10 percent of the holdings. In other words, from 90 to 97 percent of the 500 categories were not worthy of examination. There is, however, a remedy to finding significance at the category level when the data appear overwhelming. One can run the analysis with all 500 categories, export it, and create a formula to calculate the percentage of pre-19th century titles within the total holdings of each category, sort by this percentage, with a parameter of a minimum number of titles per category. After examining the results, the user can determine a cut-off. Using this method, Northwestern University Library was able to identify the fourteen categories in which 10 percent of the holdings were pre-19th century.

The library's mission determines the significance of the age of a library's collection. For some, materials may be deemed "outdated" while for others that designation will never be an appropriate description. In using the "date range" analysis to measure the age of collections, the following types of questions can be answered: Based on an analysis of the average age, what subject collections appear to be in need of updating or weeding? Have collecting patterns changed? Assuming no massive retrospective buying occurred, it is possible to trace a trend. For example, in the early decades of the 20th century, "Theater in Europe" may have been the focus of collection building in Performing Arts. What about in the last thirty years (1970–2000)? Has there been a de-emphasis on European theater and in increasing shift toward "Show Biz?" Excluding the 20th, in which century is the collection strongest? How does that compare with peer institutions? Among the oldest collections, which divisions, categories, and subjects are most prominent? In looking at subject categories as percentages of divisions, do the numbers appear to support the library's belief that the strength of the monograph collection in "Art" lies in books about painting published during the 1920s?

Analysis 3: Growth of Collections

Although the ACAS literature encourages users to analyze the "growth of the collections," this is a slightly inaccurate concept. The dates in the data are publication dates not acquisition dates, but in most cases the two are approximately the same. While it may not make sense to interpret the earliest dates in this manner, for many libraries an analysis of 20th century holdings that treats publication dates as synonymous to acquisition dates may be quite acceptable. Thus, while one might be wise to refrain from proclaiming that collecting activity dramatically increased during the 18th century in proportion to collecting in the 16th century, one may safely note if the growth rate from the decade of the 1980s was greater than during the decade of the 1990s. Any interpretation of these data, as with all data, requires some knowledge of the library's history.

"Although the ACAS literature encourages users to analyze the 'growth of the collections,' this is a slightly inaccurate concept."

Because analyses of growth sometimes rely on absolute collection sizes, the results must be interpreted in directional rather than absolute terms. For example, using the Statewide Assessment data, a comparative analysis was executed on all date ranges up to 1981, repeated on all date ranges up to 1990, and the percentage of change was calculated. The results were sorted by the percentage within each institution. This analysis informed Northwestern University Library that during the decade of the 1980s its rate of growth was moderate in comparison to other academic libraries in the state. (It ranked just above fortieth of eighty libraries.) This exercise was not difficult to repeat for the decade of the

1990s as well as a twenty-year span from 1981 to 2000. With all of these dates represented in the raw data per year, the possible combinations for analysis are many. They allow the user to speak in terms of negative and positive growth. (During the 1990s, for example, seventy of eighty Illinois libraries had negative growth rates.)

In Table 3, growth rates are calculated at the division level. Though Chemistry is one of the smaller collections within the library, the table indicates that it is proportionately one of the fastest growing (albeit in a generally slow-growth period). Did budget re-allocations or collection development policies shift ten years ago? Do the data illustrate an intended effect?

At the category and subject levels of analysis, it is again necessary to create parameters to bring sense and order to the data. For example, a comparison table of categories from four peer libraries in the Statewide Assessment Project produced forty-two single-spaced pages of data. To reduce time and effort in interpreting the results, Excel was programmed to pull only those categories with at least a 20 percent growth rate and a minimum count of 500 titles.

"Growth" can be employed to indicate if particular collections are developing as needed or in some significant way and to answer these questions: Has collection focus changed over time? While collection strengths can be measured by looking at total holdings in divisions, categories, and subjects, it is possible that these numbers portray collection development polices that no longer apply. For example, "Constitutional History and Administration" may be the largest subject category collection in the Political Science division, but the collection development policy for the last twenty years says "International Law and International Relations" should have the most emphasis. Do the data of the last twenty years increasingly reflect this shift? Which subject areas have grown the most and which the least in the decades of the last half-century? How vigorous has the library's collecting been in comparison to peers over the last twenty years? Are there categories or subfields that appear to have lost collection emphasis or been overlooked in the last few years?

Analysis 4: Title Overlap

It has often occurred that title overlap studies simultaneously demonstrate both title duplication and title uniqueness. This is not the case with ACAS. "Title overlap" is a comparison

Table 3
Ten-Year Rate of Growth within Subject Divisions

Rank	Subject Division	Acquisitions as of 1991	Acquisitions as of 2000	% Change
1	Computer Science	3,986	6,819	71.07
2	Sociology	60,967	79,936	31.11
3	Medicine	23,747	30,838	29.86
4	Chemistry	13,000	16,721	28.62
5	Law	39,537	49,798	25.95
6	Mathematics	17,694	22,033	24.52
7	Anthropology	11,742	14,384	22.50
8	Performing Arts	8,001	9,756	21.93
9	Political Science	57,820	70,305	21.59
10	Art & Architecture	61,656	74,672	21.11
11	Engineering & Technology	63,180	75,790	19.96
12	Physical Education & Recreation	3,605	4,312	19.61
13	Physical Sciences	26,468	31,570	19.28
14	Business & Economics	168,476	199,511	18.42
15	Music	24,101	28,191	16.97
16	Philosophy & Religion	79,547	92,769	16.62
17	Psychology	14,663	17,067	16.40
18	History & Auxiliary Sciences	226,592	263,739	16.39
19	Language, Linguistics, & Literature	358,387	414,517	15.66
20	Education	60,748	69,916	15.09
21	Geography & Earth Sciences	45,669	52,295	14.51
22	Agriculture	13,381	15,227	13.80
23	Library Science, Gen'l & Reference	46,881	53,125	13.32
24	Biological Sciences	26,325	29,515	12.12

of shared titles between two libraries; uniqueness is derived by comparing all institutions at once with each other, and for each institution it is a constant number. The overlap analysis table is a matrix comparing all libraries with all libraries.

"It has often occurred that title overlap studies simultaneously demonstrate both title duplication and title uniqueness. This is not the case with ACAS."

The fact that ACAS restricts overlap counts to only two libraries is a limitation. Given that in the Statewide Assessment Project there were multiple possible peer groups and each group consisted of more than two libraries, the power of this analysis was diminished. It cannot be used, for example, to study resource-sharing opportunities except within the confines of sharing with one other, single library.

Overlap was mapped by title, author, and date of publication. The overlap analysis is simple to use and comprehend. At its highest aggregate level (overall total holdings), overlap provides an indicator of the degree to which library collections are "vanilla" or identical. Such an inquiry may be of great interest, but it may also be deceptive. It is, in particular, the relatively recent rise in use of approval plans that has caused fear of over-duplication. An investigation into this phenomenon introduces another ACAS capability: it is possible to run a combination of analyses.

Both overlap and uniqueness can be cross-tabulated with date ranges. In the example above, rather than examining total holdings, a greater test of whether or not approval plans and other recent structural variables have flavored university collections vanilla can be found by running an overlap analysis limited by dates of publication. Comparisons of the last decades of the 20th century may indicate whether or not a trend toward title duplication has developed, a trend that would not otherwise be evident in a comparison of total holdings overlap.

Although restricted to matching two libraries only, overlap analysis can be utilized to answer the following questions: Which peer institution has the highest rate of duplication with the library? Excel can sort the data into an array. Chart the rates of overlap for the last ten years with each peer library on the category or subject level to see if the rate of duplication is on the rise and to map areas of duplication. In subject areas in which duplication is not desirable, are the rates too high? The results may provide evidence that assigned primary collection responsibilities need to be reconsidered. Titles that duplicate with those of a peer in any particular subject may be understood as the "core" collection and the rest (those that do not duplicate) may be construed as "sharable." Among the library's strongest collections, what percent is sharable with your peer? Which subjects have the smallest percent of core titles? When overlap is examined in terms of the age of collections, are the core collections primarily composed of newer materials? What percent of sharable titles in your strongest collections are recent acquisitions?

Analysis 5: Uniqueness

When applied to the Statewide Assessment data, "uniqueness" is a measure of unique titles in the entire State of Illinois. The frequency distribution table provides one constant figure of uniqueness per institution. To find uniqueness, items were mapped by title, author, and publication date.

The data include both the total number of unique titles per institution as well as a count of the number of libraries in possession of the same titles—e.g., "shared by 3" or "shared by 65." (It is important to not confuse these counts of shared titles with "overlap" statistics; as stated, their universe is not the same.) The data do not reveal which libraries share titles with each other. For example, Northwestern University Library has 616 titles within the "Psychology" division that are "shared by 5," but the identity of those five libraries is not provided. In some ways, this is a weakness of ACAS: it does not allow peer comparisons. In the case of the Statewide project, for example, other than geography and the fact that all the libraries are associated with academic institutions, there is no one significant characteristic that *all* of them have in common; hence, although ACAS's calculation of uniqueness might be useful for a consortium project, wherein all of the libraries are peers based on relevant characteristics, it is less useful in a statewide study among libraries with very different missions and goals.

ACAS can pinpoint unique collections at the category and subject levels. In the example in Table 4, "Show Biz" and "The Theater-General" have been identified as containing the highest proportions of unique titles. ("Parlor Magic and Tricks" is ranked second, but the total holdings are extremely low. As noted earlier, the analyst must judge the relevance of the raw figures.) The unique titles of these two categories account for almost 20 percent of the entire Performing Arts collection. Does this denote a significant collection? To answer that question, the same table can be run with peer libraries and sorted and ranked within categories.

As with title overlap, measures of uniqueness can be combined with other analyses. It is therefore possible to examine collection uniqueness within particular date ranges (age) and over time (growth). Table 4 provides a hint that crossing age within uniqueness might be interesting; perhaps the reason there is little uniqueness in the strongest subfield ("Motion Pictures") is because these are recent acquisitions. These analyses and others can answer questions regarding the richness of the collection. Although "uniqueness" is a quantified and not a qualitative value (a collection can presumably be uniquely bad), it is usually given the benefit of doubt and may be thought to indicate certain richness. If so, how rich is the library's collection? Various methodologies can be applied to answer this and each may uncover different collections of distinction. For example, which subject division has the largest size collection of unique holdings in the total collection? Which has the highest proportion when uniqueness is measured as a percentage within each division?

Table 4
Uniqueness in Performing Arts Subfields

Categories	Unique Titles	Total Holdings	% Unique within Category	% Unique within Division
Performing Arts & Show Biz	1582	3540	44.69	9.39
Parlor Magic & Tricks	13	30	43.33	0.08
The Theater- General	1443	3333	43.29	8.56
Minstrel Shows, Spectacles, Tableaux	12	71	16.90	0.07
Dancing	139	1136	12.24	0.82
Theater in Europe	56	1788	3.13	0.33
Amateur & College Theatricals	4	161	2.48	0.02
Circuses, Carnivals, Etc.	2	84	2.38	0.01
Special types of Drama	7	337	2.08	0.04
Theater- Stage, History by Period	8	416	1.92	0.05
Motion Pictures	41	3956	1.04	0.24
Broadcasting	7	697	1.00	0.04
Theater in the United States	9	1029	0.87	0.05
Theater in Asia, Africa, & Oceania	1	149	0.67	0.01
Theater in the Americas except U.S.	0	111	0.00	0.00
Jewish Theater	0	15	0.00	0.00

An exploration of the distribution of unique titles within each division is a step in discovering areas in the collection that might be worthy of special attention. Once the subject categories with significant proportions of unique titles are identified, the characteristics of these holdings can be further described in terms of age and growth. A hypothetical example follows: In terms of the total holdings of the library, the percentage of unique titles in the "Art and Architecture" division is not particularly impressive. However, in terms of the total holdings within this division, it is found that 52 percent of the titles are unique. Further analysis reveals that 87 percent of the "Painting" subject category is composed of unique titles and these unique titles are, in fact, 34 percent of all Art holdings. These percentages are much higher than those of peer libraries. Age analysis discloses that the majority of these unique titles related to painting were published in the 1920s and that no unique titles were acquired in the last thirty years. Might this 1920s collection be a target for development and retrospective buying? Or, a digitization project?

Conclusion

In the absence of a thermometer, heatstroke is a suitable indicator that the day's temperature is hot. The OCLC ACAS software is a less than perfect instrument for an analysis of monograph collections—but it is a powerful implement when used as an *indicator* of collecting patterns. As noted in this work in the form of diagnostic questions, ACAS has the potential to supply supportive statistical evidence for taking real action—e.g. updating, weeding, or re-allocating resources to exposed specializations or important hidden collections. Through cross-sectional and overtime analyses, ACAS may provide a sketch of an institution's collecting history (alone or in comparison with selected peer libraries), assess the adequacy of holdings, the effectiveness of collection development policies, and the possibility of resource sharing.

"The OCLC ACAS software is a less than perfect instrument for an analysis of monograph collections—but it is a powerful implement when used as an *indicator* of collecting patterns."

While the potential of ACAS is great, there remain serious concerns. Timeliness is one. The report of the Illinois Statewide Assessment Project, for example, is being issued in 2005 on acquisitions up to 2002. (As noted, only half of the acquisitions made 2003 were included and hence must be excluded from some analyses.) Related to this is OCLC's history of product obsolescence. With significant investment in terms of financing the purchase, mastering the skills to make use of the software, and devoting considerable time to interpret results, libraries may be concerned in regard to continuity and follow-up. If changes to collection policies, for example, are based on results an of ACAS study, the effect of those changes would want to be measured in subsequent, comparable ACAS analyses.

The most serious flaw of ACAS is the sometimes significant degree of "No Call Number Present" (essentially, missing records). This is particularly true of the older (and most unique?) titles. Titles with "No Call Number" are dropped from counts within divisions, categories, and subjects. How does one know if the missing records skew the results on these levels? Unless one has comparative data or reliable knowledge of the collection, this may be an insurmountable problem. On the other hand, the two UK consortia reported that their rates of

titles which could not be analyzed dropped after continued consultation and work with OCLC. (The final reported rates of analyzable data were 84 percent and 63 percent. Whether or not these are acceptable rates is a question individual libraries must decide.)

It may be noticed that throughout this examination many suggestions are offered for overcoming the challenges and shortcomings of ACAS tables. It may be understood, therefore, that careful and comprehensive analysis of the data is a time-intensive task. ACAS provides raw numbers, not statistics. While it is capable of generating tables of potential interest, it is up to the user to transform the distribution of raw figures into a meaningful (albeit incomplete) portrayal of collections.

Due to the unsophisticated graphing features and limited data manipulation capabilities of ACAS, it is necessary to export the results to Excel. Were one to limit analyses to the features available through ACAS, much would be missed. All of the tables as well as Fig. 3 in this report could only be produced by the exportation of multiple tables from ACAS to Excel. This can be a somewhat involved process. When faced with an ACAS table consisting of, for example, eighty rows and seventy-five columns (or much less, for that matter), it is crucial to have sufficient Excel skills that enable one to merge files, create and hide columns and rows, apply formulas, sort, set parameters, and design graphs and charts. OCLC ACAS presents a huge slab of marble from which to sculpt, with the correct tool, an impressionistic (at worst) or realist (at best) shape of monograph collections.

Notes and References

1. Mary Munroe, "Statewide Library Assessment in Illinois: A History and a Plan," PowerPoint presentation at the Annual Meeting of the Illinois Cooperative Collection Management Program (November 9, 2000).
2. *CACD Occasional News* 8 (July 1998): 2.
3. Munroe, "Statewide Library Assessment."
4. See, for example, JoAnn Jacoby, "Measurement and Analysis of Electronic Reserve Usage: Toward a new path in online library service assessment," *portal: Libraries and the Academy* 4 (April 2004): 219–32;
   Sue Samson, et al., "Networked Resources, Assessment and Collection Development," *Journal of Academic Librarianship* 30 (November 2004): 476–481;
   Irene Wormell, "Bibliometric Analysis for the Evaluation of Subject Portals," *Tidskrift for Dokumentation* 57 (2002): 21–30.
5. See, for example, Wanda Dole, "Consortium Use of the OCLC/Amigos Collection Analysis CD: The SUNY experience," *Library Resources and Technical Services* 41 (January 1997): 50–57;
   Anne C. Ciliberti, "Collection Evaluation and Academic Review: A Pilot Study Using the OCLC/Amigos Collection Analysis CD," *Library Acquisitions* 18 (Winter 1994): 431–445;
   C. Holleman, "The Study of Subject Strengths, Overlap, and National Collecting Patterns: The Uses of the OCLC/Amigo Collection Analysis CD and Alternatives to it," *Collection Management* 22 (1997): 57–69.
6. Anna H. Perrault, "The Role of WorldCat in Resources Sharing," *Collection Management* 28 (2003): 63–75.
7. CURL/RSLP Collection Mapping Project Based on OCLC/LACEY iCAS Software: Final Report (Online), www.curl.ac.uk/projects/iCAS%20final%20report.pdf (accessed September 2002).
8. Mapping Medicine: Collaborative Collection management for Biomedical Resources (Online), http://library.wellcome.ac.uk/assets/wtx017180.pdf (accessed February 2003).
9. Collection Development Committee Collection Assessment Recommendations, Preliminary Report to Orbis Council (Online), http://libweb.uoregon.edu/orbis/staffhome/CDC_coll_assess_recs-02may-draft.htm (accessed April 2002).
10. Thanks to Jeffrey Garrett, Assistant University Librarian for Collections, Northwestern University Library, for coining this description of the usefulness of ACAS.