

Information Access Through Electronic Databases For Rural Public Libraries

by Kitti Canepi

Abstract

Electronic databases allow rural libraries to expand information access beyond the limits of their material collection. How are these libraries to know which databases best meet the needs of their clientele? Research done at the onset of this project revealed no previous studies on the ability of databases to answer real questions. In this study, public library patron questions received by the Arizona State Reference Center were searched on ten different databases recommended in William Katz's *Introduction to Reference Work*, 6th edition, and available through DIALOG Online Services. Given the limitations and exploratory nature of this study, the results indicate that of the databases tested, the set of Books in Print, Magazine Database, ABI/INFORM, PAIS, and GPO Publications could qualify as a set of databases that would enable rural public library staff to find sources for 90% of the patron requests for information not found within the library.

I. INTRODUCTION AND STATEMENT OF THE PROBLEM

Discussion of the Research Question

We are at a crossroads in rural America regarding the information age; we can seize the opportunity to use technology to foster new rural economic growth independent of physical location, or be left far behind as technology rushes onward. To avoid the latter scenario will take political commitment and foresight, telecommunications availability, and "new and expanded roles for some established rural institutions," including libraries (Wilkinson 1991). The library profession is gradually shifting focus from maintaining collections to providing access to information (Malinconico 1992, Corbin 1993). In rural communities, the library is often one of the few information resources for local citizens.

Clarion University professor Bernard Vavrek, in his 1993 assessment of rural public libraries, notes the disparity between "the informational needs of rural Americans and their use of the public library to meet those needs" (Vavrek 1993). Financial resources tend to be especially limited in rural areas: every dollar spent needs to have maximum effec-

tiveness. These libraries cannot afford to build large collections of information materials; reference sources, therefore, should focus on their ability to be a gateway to the wider world of information.

Electronic databases are resources that provide improved information access. Within the space required for a single computer, a user can retrieve data from numerous sources. But this service has fees attached, and that requires choices. Libraries need to be able to evaluate databases prior to selection. They may examine reviews and product summaries for depth of coverage, update frequency, and technological requirements, but what does that tell them about the database's ability to answer real questions? How are libraries to know which databases best meet the needs of their clientele?

Statement of the Research Problem

Which set of electronic databases would enable rural public libraries to meet 90% of the patron information requests that cannot be answered within the collection?

Statement of the Hypothesis

It is hypothesized that a specific selection of databases would enable rural public library staff to find sources for 90% of the patron requests for information not found within the library.

Definition of Terms

Computer technology has had a positive influence in libraries, not only because of the ability to store large amounts of data, but also for the ability to retrieve precise bits of data from the whole more quickly than manual methods. Many library reference materials have been converted into machine-readable form and are now available electronically. Entire texts of articles and other important pieces of information have also been digitized to enable direct computer access. The two major formats for these types of electronic databases are CD-ROM and online. CD-ROM stands for Compact Disc—Read-Only Memory and consists of a 4.75-inch disk, similar to a music CD disc, that can hold up to 250,000 pages of information (Katz 1992). Both indexes and basic reference works (directories, encyclopedias and handbooks) are available in this format. Some of these are CD-I, or Compact Disc Interactive, which include video images, and some are Multimedia, meaning they include both images and sound. Online means that the database is residing on a large computer somewhere which the user accesses over a telecommunications link. Retrieval of information is faster with this format, since you are taking advantage of the larger

computer's capabilities, and the information tends to be more current since the database owner can update it as often as they like at less expense than remastering CD discs. A disadvantage to online databases is that you generally pay per each minute or hour of connect time. This may become less of a factor as more databases become available over the Internet.

Information access relates to the ability to find information outside the physical limits of the local library. This includes finding out what information is available, where it is located, and how to get it. This ability is very important in rural communities where distance can cause isolation, where the library is one of few resources available, and where limited funds result in small collections of print materials. This study focuses on the needs of rural libraries. "Rural" is defined by the Bureau of the Census as "places of less than 2,500" in population and places outside incorporated and census designated places, including the rural portions of extended cities (1990 Census of Population and Housing 1993). The main function of this definition is to distinguish rural areas from "urban" areas. To further separate urban and rural territory in the vicinity of large places, the Census Bureau uses the term "urbanized area", which is comprised of a central place and surrounding densely populated territory (at least 1,000 people per square mile) that together have a minimum population of 50,000 persons (1990 Census 1993). Between these two Census figures, the Center for the Study of Rural Librarianship uses "non-metropolitan" places (i.e. outside a metropolitan area) with a population of less than 25,000 to define the term rural (Vavrek 1993). This definition is also used by the American Library Association (Mumma 1991). By this definition, approximately 80% of the public libraries in the United States serve rural populations (Cox 1993).

Underlying Assumptions

One of the assumptions made for conducting this research is that Arizona is primarily a rural state. The 1990 Census lists Arizona as having a total population of 3,665,000; 87.5% of that population live in a small number of urbanized areas. 2.2 million out of the 3.6 million total population live in the metropolitan area of Phoenix-Mesa, 667,000 live in Tucson, and the remainder are spread out over approximately 130,000 square miles. Other western states generally considered rural show similar statistics: 82.4% of Colorado's population live in urban areas concentrated mostly in Denver-Boulder; 73% of New Mexico's population live in urban areas, mainly in and around Albuquerque; 87% of Utah's population live in the urban Salt Lake-Ogden area; and 88.3%

of Nevada's population live in urban areas, divided between Las Vegas and Reno. All of these states have a small number of urban areas which account for the majority of the population count, with the remainder of the population spread out over vast territory.

Another assumption made for this study is that the questions asked by library patrons in a given period of time are representative of the kinds of questions normally asked on an ongoing basis. Therefore, databases that can provide answers to randomly chosen sample questions from a given year should do equally well at providing answers for future questions. A third assumption relies on the integrity of volunteer study participants; that they will perform to the best of their ability, that their methods are appropriate, and that their results are legitimate.

Limits of the Research

The databases chosen for this study were limited to a select group of reference and citation databases recommended in William Katz's *Introduction to Reference Work*, 6th edition and also available through DIALOG Online Services. The rationale for this was that a public library considering first time access to electronic databases would likely depend on a recognized authority and an established service. This study was designed to be exploratory, rather than exhaustive, which called for limitations to the number of databases studied. By using a single service to access all of the tested databases, a common search interface could also be maintained. Ten databases were eventually selected based on these two criteria. It was not considered pertinent to this study to ensure that all subject areas were covered among the selected databases, nor to do stratified random sampling of the study questions by subject area. The purpose of this study was to mirror rural library situations where patron questions are unpredictable and not likely to cover all subjects. Recommendation and availability were the primary criteria for database selection, and simple random sampling was used to pull the sample questions.

Another limit to this project is that volunteers were used to do the actual database searches without controlling for their skill level or search methods. Being able to complete the project was given priority over strict control for every possible variable. The search results were checked for validity, however, through a separate quality control check of random question and database combinations by two more experienced searchers. All of the participants in this study were graduate Library Science students at the University of Arizona at the time of the study. While the majority of them were not experienced in using DIA-

LOG, they all were experienced with information searching and with the use of electronic resources. This reflects actual conditions in rural public libraries where database experts are rarely available. The use of experts would have lent a greater air of objectivity closer to that of a lab study than a field study. It is also recognized that the ideal way to verify the results of this study would have been to replicate it entirely by two different search groups who would record their strategy, compare results, and then repeat the process. Because such measures were not done here, it is suggested that this be viewed as a pilot or exploratory study.

Significance of the Research

Rural libraries need to gauge the effective and efficient use of their resources even while exploring the use of technology in providing information access to their community. Perhaps the most significant and unique aspect of this study is that it is aimed at demonstrating the practical, versus theoretical, use of electronic databases. While reviews are often available, libraries still have only trial and error to determine whether a recommended purchase was indeed the appropriate one for their type and size library. Since libraries have limited funds that force them to be very selective in their purchasing decisions, they would benefit greatly from learning the results of someone else's trial and error, or from having an available source that tells them which databases are best suited to their type of library. Trial and error experiences are often shared through conferences and publications. A source that lists which databases are best suited to which type of library does not currently exist.

II. LITERATURE REVIEW

Review of Relevant Literature

A joint congressional hearing in 1982 addressed "The Role of Libraries and Information Technology" in regards to rural America (Joint Congressional Hearing 1982). Oral testimony included a number of statements concerning the need for libraries to be "the focal point for all the information needed by the community," since it is not economically feasible for every rural resident to have direct access to all of the existing databases and information systems (Margaret Warden, former Montana Senator, Joint Congressional Hearing 1982). In Kenneth Wilkinson's article "Information Access in Rural Areas" (1991), he points out that "[d]istance is perhaps the most enduring characteristic of the quality we call 'rural,' and distance impedes access." By creating a community information access point through the local library, the limi-

tations of distance can be reduced. Library staff can also serve as "the human link between the learner and the equipment as we adopt more and more technology" (Dr. Betsy Peters, University of Wyoming, Joint Congressional Hearing 1982). The rural library clearly is "a link in the resources of the community" (Dr. Bernard Vavrek, Clarion University, Joint Congressional Hearing 1982).

Librarians, as well as non-librarians, have written about the need for libraries to readjust their methods and views to incorporate changing technologies. It has been stated that both librarians and society as a whole "want to perceive the library as an information center" (Vavrek 1990). Discussion of the National Information Infrastructure emphasizes the danger of creating classes of information "haves" and "have nots." Many people in government and the library field hope to take advantage of the library's role as "a vehicle for democracy in our society" to bridge that gap (McCormick 1994). Given sufficient governmental support and funding, "public libraries can provide access to the masses of information for society's have-nots at costs they can afford" (Goddard 1994).

Several state governments have started putting this philosophy into practice. The initial plan for the North Carolina Information Highway, a statewide fiberoptic network, specifically designated libraries as public access sites (Rogers 1994). New York undertook a special pilot project involving connecting rural libraries to the Internet, based on the concept that "if rural librarians were given the tools and training to use networked electronic information, they could do so effectively and thereby improve the quality of service they offered their patrons" (McClure, Babcock and Nelson 1994). The project was rated an overall success as information resources were greatly expanded, local communities started viewing the library as a "vital information center," and less patrons needed to be referred to larger libraries.

A study done by Polly Mumma in 1990, demonstrated that "most rural libraries do not generally have the same technological advances at their disposal as their urban counterparts" (Mumma 1991). In 1993, Bernard Vavrek found that 48% of rural populations would like their libraries to provide computerized information, yet only 6% of the libraries actually did (Vavrek 1993). A study of rural Nebraska libraries in 1993 made no mention of computer technology as being available in any of the surveyed libraries (Cox 1993). And yet, a study done in 1989 indicates that rural residents are no less likely to accept and use computer technology than their urban counterparts (La Rose and Mettler 1989).

Access can refer to helping patrons find citations and locate information, or having physical access to a building (including ADA concerns), or delivering physical information containers such as books and documents (McCann 1994). While there needs to be a balance between collections (the physical materials on hand) and access (the ability to locate physical materials housed somewhere else), this study is primarily concerned with expanding the information access ability of the rural library. Limited funding will always restrict the amount of materials on hand, but by providing the means by which citizens find out what other information is available, where it is located, and how they can get it, libraries allow users to overcome collection limitations. This is the significance of a "virtual" library; not a library without walls, which implies no need for a physical building, but a library beyond walls that can reach out to a larger clientele, or bring the world of information into any library (Crawford 1994).

Acquisition of the electronic resources that expand information access can be a complicated and frustrating experience, especially for those new to these formats. How is a selector to evaluate the product? In 1992, Cheryl LaGuardia and Stella Bentley looked at the need to expand collection development policies to include electronic databases. They suggested that set criteria should be established to address content, relevance, usefulness, cost, and accessibility. To judge the content of a database, selectors with subject expertise should be involved, armed with database reviews and a set of technology-based criteria that include "site-specific technological requirements and media-specific evaluative measures" (LaGuardia 1992). An Iowa State University study that same year compared electronic periodical indexes offered by three different companies. The evaluation criteria they used was from Katz's *Introduction to Reference Work*: scope, duplications and gaps, depth of indexing, timeliness, format, and subject headings. The assigned task force selected databases that "in general" fulfilled the research needs of undergraduate students, and found that each product had strengths and weaknesses, so that no one vendor would alone meet all of their needs (Fry 1994). This conclusion has since been echoed by Carol Tenopir: "no one medium or set of sources can satisfy all requests or all users" (Tenopir 1995). The *Guide to Selecting and Acquiring CD-ROMs, Software, and Other Electronic Publications* lists specific policy, service, technical, and cost considerations that should be addressed by an electronic format selection criteria. The authors consider user needs as a "prime consideration," and suggest defining local needs by subject areas needing specific products in electronic format, patron and staff groups that will benefit from these products, the ease of

use and depth of information appropriate for the intended groups, comparison of the intended products with other electronic or print products, and location issues (Bosch 1994). As stated by Chris Armstrong in an address to the Scottish Library Association in 1994, the need to evaluate databases is "inescapable"; yet currently "users have no fixed criteria by which to judge, and suppliers have no standards by which they can measure" (Armstrong 1994). The Library Association in the United Kingdom has set up a Centre for Information Quality Management for its members to report database problems. The Centre will forward the problem on to the appropriate body (information provider, online host, etc.) and route responses back to the user. By serving as a clearinghouse, they hope to gather statistics and "anecdotal evidence" to demonstrate the need for quality maintenance by database providers. Their long-term goal is to create "a series of measurements by which databases can be assessed" (Armstrong 1994). All of these efforts help towards establishing general criteria for evaluating electronic resources. But what about the question of which databases are best suited to the needs of a specific clientele?

Most recent to the conclusion of this study, Mick O'Leary of ONLINE Magazine looked at this problem. His approach was to create an Advisory Board of eight online subject experts to help select the one hundred "most significant online databases in every branch of human knowledge and endeavor," resulting in the book *The ONLINE 100: ONLINE Magazine's Field Guide to the 100 Most Important Online Databases* (O'Leary 1995). O'Leary concedes that some subjects are better represented in the book than others, reflecting the emphasis they receive from online users who tend to favor business and science-technical information. Preference was given to source versus bibliographic databases, and full-text or abstract versus citation only, in the view that "[a]n online database should be an end in itself, rather than one step in a laborious process" (O'Leary 1995). In this most database users would likely agree, although in some cases a citation might be preferable to no information at all. Most of the databases O'Leary selected are available on mainstream online services, both professional and consumer oriented. The databases in *The ONLINE 100* are grouped into ten broad subject categories and each is described by content, search points, what not to use the database for, unexpected or generally unknown facts, and key facts such as type of database (bibliographic, fulltext, etc.), span of coverage, producer, and how to access the database (online hosts, Internet, CD-ROM, costs). In reviewing the one hundred databases chosen, only one database used for this current study, *Bibliography Master Index*, was not included, although the GPO

Publications Monthly Catalog received only an honorable mention. Had The ONLINE 100 been available prior to the onset of this study, it may have proved a good source for the selection of the databases.

Summary and How Current Study Relates

The literature indicates that technology is an important concern and that libraries should be access points for rural patrons. How then are rural libraries to provide the necessary access for their users? Telecommunication technologies could be used to equalize access to electronic resources, with libraries as the logical intermediaries between the technology and the rural user (Senkevitch 1994). Although the limited funds available in rural libraries is seen as a prohibition against developing electronic access, it is also one of the primary reasons for doing so: if you can only afford a limited amount of materials, you can augment that by providing access to information outside of your own collection. To do that, however, you need to know which set of databases will answer the largest portion of the questions you are likely to receive in your library.

Up until the publication of O'Leary's study, no one had undertaken to compose a single list of recommended databases. Reviews of electronic databases were scattered and tended to focus on ease of use or depth of coverage. Research done at the onset of this project revealed no previous studies on the ability of databases to answer real questions. Certainly the issue of appropriateness for rural library populations has never fully been addressed. As stated by Ameritech vice-president Greg Brown, libraries need to identify the applications that most benefit their patrons (Brown 1995). This study should begin the process of filling in this gap in librarian's knowledge of electronic databases. In the increasing glut of information resources, it is time for availability to give ground to effectiveness.

III. METHODOLOGY

Research Design

This study used the expert opinion contained in William Katz's *Introduction to Reference Work*, Volume I, 6th Edition, to compile a list of recommended electronic databases. This work is frequently used as a text in Library and Information Science graduate-level courses and is widely recognized in the library field. Databases were first chosen based on their applicability to public libraries, then reduced to those also available through the DIALOG Online Service. The chosen databases were, in alphabetical order, ABI/INFORM, Biography Master

Index, Books in Print, ERIC, GPO Publications, Magazine Database, MEDLINE, Newspaper Abstracts, PAIS, and PROMPT (see Appendix A for brief descriptions).

A list of questions from the Arizona State Reference Center served as the representative population of questions. The State Reference Center is a free service provided by the Department of Library, Archives and Public Records (State Library). Public libraries in Arizona can submit questions to the State Reference Center after they have exhausted the sources within their own libraries. Since the State Reference Center does not charge for its services, there is every reason to believe that for the time period of the representative questions, libraries sent in all of the questions they could not answer with materials in their own library, rather than only sending the more difficult questions. A list of actual questions relayed to the Reference Center should, therefore, be a good indicator of the type of questions libraries in Arizona receive from patrons. Many of the contributing libraries were fairly small with limited collections, and the majority of the libraries did not have access to electronic databases at the time they sent in the questions used for the study population, although many have since established access to OCLC's First Search through the efforts of the State Library Extension Division.

A simple randomly selected sample of the Reference Center questions was tested against the selected list of databases. Volunteers from the University of Arizona School of Library Science were each given a list of the sample questions and assigned one of the selected databases to search. Answers were noted on a data collection sheet in the space provided following each question. Questions for which no answer was found were left blank on the data collection sheet. The results were analyzed by number and percent of "hits" per database.

Potential Threats to Validity

It is recognized that this is primarily a descriptive study and it is hoped that this study will be considered a valid exploratory effort in an area where little research work has been done. Maintaining controls such as strict simple randomization of question selection, and searching each question in each database ensured the internal validity of the study. In addition, the quality control check done once the study was completed re-tested a randomly selected group of questions and databases representing 10% of the total question and database combinations. A comparison was done to answers found in the original study to verify the interrater reliability of the initial search results.

Since Arizona is a predominantly rural state, a simple random sample of questions sent to the Arizona State Reference Center should give a fair picture of the kinds of questions asked by patrons in rural Arizona libraries, and therefore make these results useable by Arizona libraries. While the results are likely generalizable to other rural libraries, it is left up to libraries outside of Arizona to make that determination for themselves. Again, this study makes no claims to be more than exploratory in nature.

Study Sample

This study used a simple random sample of questions received by the Arizona State Reference Center. A population of 1045 questions were used for this study; all were received by the State Reference Center between October 1989 and May 1990. Shortly after that time period the State Library began the project to help Arizona public libraries establish online database service through OCLC's First Search. Questions received before the majority of the libraries had any electronic information access was considered more representative of the type of questions asked in Arizona public libraries. These older questions had been stored in a database using different software than was currently being used, causing extraction problems that limited the time period covered and excluded questions from January 1990. The identity of the source library was stripped from the questions before downloading.

A modified Scheffe' technique was used to determine sample size for a level of significance of .05. A random sample of fifty-two questions was determined to be of sufficient size to account for both Type I and Type II statistical errors. The questions had been assigned sequential numbers by the State Reference Center that appeared to be related to the date received. Because of gaps in the numbering sequence, new sequential numbers were assigned to the questions and a computer spreadsheet random number generator function was used to select the question numbers. A data collection sheet was created listing the selected questions and providing space to record the search results (see Appendix B).

Data Collection

Each volunteer searcher was assigned a database and given a data collection sheet of all fifty-two questions. Some of the volunteers searched an entire set of fifty-two questions in a single database, while others searched only a partial list of five to twenty-five questions. The number of questions received was determined by the volunteer, given the amount of time he or she expected to contribute to the project.

Divided among the volunteer searchers, each set of fifty-two questions was searched on each of the ten databases; there was no overlap in coverage. Searchers with no previous experience using the DIALOG Online Service were provided with a brief training session. To help level out the varied experience levels, those volunteers with previous DIALOG experience were assigned databases with which they were not familiar. Since the volunteers were all graduate students at the University of Arizona School of Library Science (Library School), access to DIALOG was made available under a previously received Library School grant. Searchers were provided with logins and passwords, were instructed how to access DIALOG from any computer with telnet telecommunications capabilities, and were left on their own to arrange when and where they would conduct the searches. Many made use of the computer lab at the Library School. The bulk of the searches were completed between July 17 and August 3, 1995; some sets of questions were completed shortly thereafter. Out of twenty-six students who volunteered for the study, seventeen completed their assigned searches. The remainder of the searches were completed by the author following the initial study period.

The interrater agreement was for the volunteer to search each given question in the assigned database and record the results on the data collection sheet whenever an answer was found. A citation to a source covering the same subject as that of the question was considered a legitimate answer. It is recognized, however, that a citation and an answer are not necessarily the same thing in cases where a citation has not been specifically requested. Customer dissatisfaction is possible if the desired citation item cannot be obtained by the local library through interlibrary loan or document delivery. For the purposes of this study, however, the broadest definition of information access was used: finding out what information is available, where it is located, and how to get it. Under that definition, citations were considered acceptable.

The searchers were not instructed to record their search strategy, although some did. They were instructed to write the answer on the data collection sheet when they found it, although a few did not, merely marking "yes" or "no". This was still considered valid as the categories of data are mutually exclusive, i.e., the answer was either found in the database or it was not. Since there was no intrinsic order to the categories, only nominal measurement could be used to analyze the data. A spreadsheet was set up to record the results, with columns representing the databases and rows representing the questions. A number "1" was entered into the corresponding cell when an answer was found; a num-

ber "0" was entered was an answer was not found. In this way, total hit rate could be easily tabulated. The hit rate was then divided by the total number of questions to create a percentage.

Reliability of Data Collection Instrument

In order to verify the validity of the data collected, a quality control check was done by two other library science students with greater expertise in DIALOG searching than the original group of volunteers. This was done to determine whether an answer was not found due to searcher error rather than because the answer was not there. A 10% random sample of the total database and question combinations (fifty-two out of five hundred twenty possibilities) was used. The questions and databases were selected through use of a spreadsheet random number generator, then divided between two data collection sheets. The sheets were blindly drawn to give to one of the two quality control searchers. The results of the check generally paralleled that of the original study, although there were a few differences that could be contributed to variations in subject knowledge, or question interpretation, that could be expected from the human element involved in this type of study. Individual searchers bring their own experiences, knowledge, and creativity to a search, resulting in a variety of possible approaches to a question. In only three instances did the quality control searchers find at least partial answers where they had not been found previously; in three other instances, the searchers did not find answers that had previously been found.

One of the limitations to the search procedure, that became increasingly obvious in the quality control check, was the format of the questions themselves. The original questions received by the State Reference Center were entered into the computer in abbreviated form that sometimes left the exact nature of the question unclear. In some cases the spelling of words was incorrect, although it is not known whether due to misspelling in the original questions received by the Data Center or due to typing error. The questions were modified only slightly for consistency of language and form when entered on the data collection sheet. The brevity of the questions allowed for no analysis of original intent, therefore every attempt was made to retain the ambiguity on the data collection sheet so as not to bias the results, excepting the bias of the searcher in creating the search strategy. It is hoped that in an actual library setting, a database searcher would request further reference interviewing to better define the nature of the information sought. Misspellings were also not corrected on the data collection sheet, but were noted with a "[sic]" wherever easily identified. Again,

since the source of the misspelling was not known, and searching of "real" questions was a key to this project, every effort was made to not bias or alter the authenticity of the study.

IV. STUDY RESULTS

Analysis Procedure

The main objective of the analysis was to rank the selected group of databases on their ability to answer the largest number of sample questions. The total number of hits per database were tabulated and a hit rate percentage figured based on the ratio of answers found in each database per the total number of questions in the sample. A list of databases was then created, in order by total hit ratio, indicating total number of hits and percentage of hits per total number of questions (see Appendix C).

Analysis of the Data

At the outset of this study, it was not supposed that any database would emerge as a single source, but rather, it was hoped that a limited group of databases would be determined to be desirable. In general, the total number of hits per database were lower than expected. One database did emerge as having a greater superiority of hits, based on the given criteria of acceptable answers: Books in Print. A full 65.38% of the questions were answered by citations in this database, since titles of books about the same subject as the question were considered to be answers. One database came up with an extremely low number of hits: Biography Master Index. Only 1.92% of the questions found answers or citations in this database. The remaining databases had hit rate percentages of between 13.46% and 32.69%, with three of them at 19.23%. Interestingly, the database with the lowest showing was also the only database not listed in The ONLINE 100. This could be construed as confirming O'Leary's choice in excluding this from the databases that most libraries should consider among their first choices.

Four databases emerged from this study as having a 25% or greater hit rate: the previously mentioned Books in Print (BIP), Magazine Database (MagazineDB), ABI/INFORM, and PAIS. Because of the determination that a citation would be considered a legitimate answer, it was not surprising that BIP had the greatest number of hits at 65.38%: BIP lists all books currently in print, soon to be published, or recently out of print for the entire U.S. book market, without restriction as to subject. MagazineDB emerged with a 32.69% hit ratio, the next highest after BIP. The wide coverage of periodicals in all subject areas could

account for some of this success. ABI/INFORM was next in line with 28.85%, and PAIS came in fourth at 25%. It is interesting to note that business (ABI) and public affairs (PAIS) databases ranked so highly. This may indicate both something about those databases (that they are fairly broad in scope), and something about the sample questions (that a great number of them are related to business and public affairs issues). Both BIP and MagazineDB are more general in nature, as is the database closest below 25%, Newspaper Abstracts. The top four databases seem to be good candidates for answering the type of questions one might expect to be asked in rural public libraries, based on their hit rate. While this study is at most preliminary and exploratory, the emergence of ABI/INFORM and PAIS databases does seem significant.

V. CONCLUSIONS AND SUGGESTIONS FOR FURTHER STUDY

Conclusions

The original research problem was which set of electronic databases would enable rural public libraries to meet 90% of the patron information requests that cannot be answered within the collection. The data collection results were analyzed to determine the smallest set of databases, out of the ten studied, that would provide the greatest total coverage of the representative questions. By looking at the overlap of answers among the four databases with the highest hit rates, forty-six out of fifty-two, or 88.46% of the sample questions could be answered within this set of databases. This does assume that one is willing to accept a citation as an answer, although it could be argued that it is only a partial answer. Material access does play a role in customer satisfaction, although it is not an issue addressed in this study. Additionally, each of these four databases — BIP, MagazineDB, ABI/INFORM, and PAIS — answered at least one question not answered by any other database.

Of the six questions not answered by any of these four databases, five questions were not answered in any of the databases studied. This means that a maximum of 90.38% coverage could be achieved by any grouping of these databases. The one other question not covered by the top four databases was answered in both Biography Master Index and GPO Publications (GPO). Given the overall poor showing of Biography Master Index (1.92%) in comparison to GPO (13.46%), GPO would appear to be a better choice between the two for answering that one additional question. By adding GPO to the other four databases, a 90.38% hit rate can be achieved. No other grouping out of these ten

databases would produce the same hit rate (excepting the previously mentioned Biography Master Index). Given the limitations of this study, the results indicate that of the databases tested, BIP, MagazineDB, ABI/INFORM, PAIS, and GPO could qualify as at least one of the sets of databases that would enable rural public library staff to find sources for 90% of the patron requests for information not found within the library.

Suggestions for Further Study

Electronic databases are one viable option for libraries seeking to expand information access beyond their walls and material collection. When funding is limited, access costs become increasingly important. This study looked at online access through a single provider. The range of DIALOG hourly charges for the selected group of databases necessary to reach a 90% hit ratio is \$35 to \$114 an hour (see Appendix A). Limiting search time helps contain those costs within an affordable range, which then becomes an issue of searcher experience and training. Desired percentage of coverage has to be weighed against what the library can afford. Dropping ABI/INFORM, the most expensive database, from the recommended set would reduce the overall cost while retaining an 88.46% hit ratio, as ABI/INFORM only had one answer that was unique. MagazineDB (\$84 per hour) also only had one unique answer; dropping it from the group would result in an 86.54% coverage rate. It must be remembered, however, that BIP, the database upon which these high hit ratios are based, is a citation-only database, while ABI/INFORM and MagazineDB both include abstracts and some full-text. Material access may become an issue for consideration. Further study could be done to compare costs of various online services, as well as CD-ROM versions, for access to these databases. As more databases become accessible over the Internet, it is possible that costs will decrease.

The main purpose of this study was to test the methodology of evaluating databases by their ability to answer real questions asked in a specific type of library. It is hoped that this will inspire others to adopt similar approaches when examining information resources. Unless librarians start approaching the influx of information resources in terms of practical use, we will not only become overrun with choices, but likely waste funds in the trial-and-error of determining what works and what does not. Further studies like this one are required that deal with different sets of sample questions and different sets of databases. The databases listed in The ONLINE 100, as well as those available on OCLC's First Search, seem good candidates for another study. Above

all, it is time for rural public libraries to make use of technology to expand information access, and it is also time for librarians to start analyzing the appropriateness of available information for the people whom they serve.

Kitti Canepi is currently employed with the Extended Campus Service Libraries of the East Tennessee State University in Kingsport, Tennessee.

APPENDIX A - DATABASE DESCRIPTIONS

Information from William A. Katz's Introduction to Reference Work, Volume 1: Basic Information Sources, 6th ed., NY: McGraw-Hill, 1992. All also appear in DIALOG Database Catalogue, Knight-Ridder Information, Inc., 1995.

1. Books in Print-DIALOG file 470, 1979 to present, includes Subject Guide, Supplement, Forthcoming, Out of Print, Paperbound, Children's with Subject, Scientific & Technical Books, Medical Books, Business & Economic, Religious, \$65 per hr.
2. Magazine Database (Infotrac Magazine Index)-DIALOG file 47, covers 435 periodicals with annotations, includes book reviews, \$84 per hr.
3. ABI/INFORM-DIALOG file 15, international coverage, detailed abstracts, covers 800 titles, including most aspects of business and the economic world, \$114 per hr.
4. PROMPT (Predicasts Overview of Materials and Technologies)- DIALOG file 16, worldwide view of business and individual companies, covers 1200+ titles, \$126 per hr.
5. Newspaper Abstracts-DIALOG file 603, covers 25 newspapers including New York Times, Chicago Tribune, LA Times, Wall St. Journal, detailed abstracts, \$84 per hr.
6. Biography Master Index-DIALOG file 287, includes names from over 700 sources, lists essential facts, \$63 per hr.
7. GPO Publications (U. S. Government Printing Office Catalog)-DIALOG file 166, lists current government publications, \$35 per hr.
8. PAIS International-DIALOG file 49, government, legislation, economics, sociology and political science, covers 1400 journals and 6000 other items, \$75 per hr.
9. ERIC (U. S. Educational Resources Information Center)-DIALOG file 1, index and abstract of reports and other materials on education and related fields, includes library science and social sciences, \$30 per hr.
10. MEDLINE-DIALOG file 155 (1966 - present), 154 (1985 - present), comprehensive medical index, citations with abstracts from 3400 journals in the U. S. and 70 some other countries, includes related fields in social sciences, technology, agriculture, \$36 per hr.

**APPENDIX B—STUDY QUESTIONS/DATA
COLLECTION SHEET**

Study of Electronic Databases on DIALOG
(7/17 - 8/3, University of Arizona)

Name of Database _____

List either the answer or the bibliographic citation of a source which might answer the following questions:

1. What was the Consumer Price Index for Oakland, CA for the following dates: 12/31/87 _____ , 12/31/88 _____ ,
12/31/89 _____

2. A source for the following auto repair manuals:
1977 Ford LTD 425cc _____
1976 Chevrolet Monte Carlo 325cc _____

3. A source for determining the value of a book

4. The value of silver coins minted by the Franklin Mint in 1976

5. General information on 6 engineering companies

6. A source for Stephen King's Wizard and Glass and Wasteland

7. List attorneys in La Puente, CA

8. Information on the care & feeding of a Prarie [sic] Falcon

9. Information on the Department of Defense (history, mission, etc.)

10. The source of a book or information on Myotonic Dystrophy

11. Information on farming and ranching in Alabama

12. Is there an AZ business called "S&H Pipeline Construction"?

13. Who adapted Victor Hugo's Les Miserables into play?

14. Source for a book with pictures of "Ocotilla & 7 Angels" and "Saharo & 3 Angels" by De Grazia

15. List the names and addresses of companies that do conversions of large buses to vans for entertainers

16. What Federal Regulation governs nuclear power plant workers?

17. What is the definition of "crystal trylon"?

18. A source for copies of Department of Economic Security (DES) programs

19. A source for a songbook on artist Emmylou Harris

20. Information on wild muskrats

21. Information on mountain man Pauline Weaver

22. A source for a book on a Globe-Roosevelt, Arizona, country doctor

23. What are the average fees of visiting nurses in Tucson and Phoenix, AZ?

24. Information on a horse race run on March 17, 1957

25. What is the address and phone number of the Gregor Valor Mutual Fund?

26. Albany & Long Island, NY, yellow page listings for MASSEY

27. Information on Scream Therapy

28. Who is the author of the poem that contains the lines "Burn us and we will rise again...?"

29. A source for a book on puberty written for pre-teens

30. What is the poem or saying about an acorn growing into an oak tree, and who wrote it?

31. A source for the Assayer's Handbook

32. What Chinese poet wrote the poem that contains the lines "Oh, Nefarious War! I see why arms..."? _____

33. Information on the history of antique trucks

34. What are the instructions for making a cradleboard for an Indian papoose? _____

35. What is the address and phone number of the Ruby Lamp Company? _____

36. What U.S. Code regulates Collection Agency procedures?

37. What was the value of the Oppenheimer High Yield Mutual Fund on February 15, 1979? _____

38. A source for the book Hannegan Meadow _____

39. Information on the artist Tomie Fusiurto _____

40. Instructions on how to restore books that have been sprayed with fire extinguisher foam _____

41. What is the coat of arms for the Serrano family? _____

42. What is the definition of the phrase "san pa ku"? _____

43. Information on forest fires _____

44. A source for a book on the values of old books _____

45. Information on a wind or windstorm known as "haboob" found in Egypt, India & Arizona _____

46. What are the words to the song "Abraham, Martin & John"?

47. Information on building and stocking a pond, including types of fish to stock it with _____

48. Information on artist Aurilio Yammarino _____

49. A source of costume designs for showgirls from the Ziegfel [sic] Follies _____

50. Information on the history of education, along with current qualifications and average earnings

51. Information on the construction of bridges

52. Information on how to make baskets using rags

APPENDIX C- DATABASE RESULTS LIST

Database (in order by hit rate)	Total # of Hits	% of Total Questions
Books in Print	34	65.38%
MagazineDB	17	32.69%
ABI/INFORM	15	28.85%
PAIS	13	25.00%
Newspaper Abstract	12	23.08%
ERIC*	10	19.23%
MEDLINE*	10	19.23%
PROMPT*	10	19.23%
GPO	7	13.46%
Biography Master Index	1	1.92%

* Databases listed alphabetically since hit rates were equal.

BIBLIOGRAPHY

- Armstrong, Chris J. "Database Evaluation and Database Quality." *Proceedings of the Annual Conference—Scottish Library Association*, issue 80 (1994): 75-79.
- Bocher, Robert. "Access and Cost Issues in Rural Networking." *Internetworking Rural Libraries Institute (UW-Milwaukee - SLIS): State of Wisconsin Department of Public Instruction*, 1994.
- Bosch, Stephen, Patricia Promis, and Chris Sugnet. *Guide to Selecting and Acquiring CD-ROMs, Software, and Other Electronic Publications*. Chicago: American Library Association, 1994.
- Brown, Greg. "The Management of Knowledge: Libraries Can Transform Communities." *American Libraries*, vol. 26, no. 3 (March 1995): 222-224.
- Corbin, John. "Competencies for Electronic Information Services." *The Public-Access Computer Systems Review*, vol. 4, no. 6 (1993): 5-22.
- Cox, Jamie Suzane. "Services Offered in Rural Libraries in Nebraska." *Rural Libraries*, vol. 13, no. 1 (1993): 13-34.
- Crawford, Walt. "The Virtual Library: Virtually Impossible." *AMIGOS Fall Conference*, November 1993.
- Fry, Sally A., and Kathy A. Parsons. "Comparative Analysis of IAC, UMI, and Wilson Database Tape Products." *The Serials Librarian*, vol. 25, no. 1-2 (1994): 133-34.
- Goddard, Stephen B. "The Information Superhighway: Crisis & Opportunity." *Library Journal*, vol. 119, no. 12 (July 1994): 56.
- Joint Congressional Hearing on the Changing Information Needs of Rural America: The Role of Libraries and Information Technology*, July 21, 1982. U.S. Department of Agriculture and National Commission on Libraries and Information Science. U.S. Government Printing Office, 1984.
- Katz, William A. *Introduction to Reference Work*, 6th edition. McGraw-Hill, 1992.
- LaGuardia, Cheryl and Stella Bentley. "Electronic Databases: Will Old Collection Development Policies Still Work?" *Online*, vol. 16, no. 4 (July 1992): 60-63.
- LaRose, Robert and Jennifer Mettler. "Who Uses Information Technologies in Rural America?" *Journal of Communication*, vol. 39, no. 3 (Summer 1989): 48-60.
- Malinconico, S. Michael. "Information's Brave New World." *Library Journal*, vol. 117, no. 8 (May 1992): 36-40.

- McCann, Jett. "Why Document Delivery Services Should be Handled by Serials Librarians." *At Your Service*, no. 30: 6-9, 20.
- McClure, Charles R., Waldo C. Babcock, and Karen A. Nelson. *The Project GAIN Report: Connecting Rural Public Libraries to the Internet*. Project evaluation report prepared for NYSERNet. Manlius, NY: Information Management Consultant Services, 1994.
- McCormick, Edith J. "President-elect: 'Libraries Guarantee Equity on the Information Highway.'" *American Libraries*, vol. 25, no. 7 (July/August 1994): 688-689.
- Mumma, Polly S. "Technology and the Rural Library." *Rural Libraries*, vol. 11, no. 2 (1991): 7-20.
- 1990 Census of Population and Housing - Guide; Part B. Glossary*. U.S. Department of Commerce, Bureau of the Census. Washington, D.C.: U.S. Government Printing Office, 1993.
- O'Leary, Mick. *The ONLINE 100: ONLINE Magazine's Field Guide to the 100 Most Important Online Databases*. Wilton, CT: Pemberton Press Books, 1995.
- Rogers, Michael. "NC Builds Statewide Fiberoptic Information Network." *Library Journal*, vol. 119, no. 12 (July 1994): 29.
- Rooks, Dana. "The Virtual Library: Pitfalls, Promises, and Potential." *The Public-Access Computer Systems Review*, vol. 4, no. 5 (1993): 22-29.
- Senkevitch, Judith J. and Dietmar Wolfram. "Equalizing Access to Electronic Networked Resources: A Model for Rural Libraries in the United States." *Library Trends*, vol. 42, no. 4 (1994): 696-710
- Tenopir, Carol. "Online Databases: Integrating Electronic Reference." *Library Journal*, vol. 120, no. 6 (April 1995): 39-40.
- Vavrek, Bernard. *Assessing the Information Needs of Rural Americans*. Clarion, PA: Clarion University, 1990.
- Vavrek, Bernard. *Assessing the Role of the Rural Public Library*. Clarion, PA: Clarion University, 1993.
- Wilkinson, Kenneth P. "Information Access in Rural Areas." *Rural Libraries*, vol. 11, no. 1 (1991): 53-65.