ACCESS TO THIRD WORLD SCIENCE IN INTERNATIONAL SCIENTIFIC AND TECHNICAL BIBLIOGRAPHIC DATABASES

Gretchen WHITNEY
Graduate Library School, The University of Arizona
1515 East First Street, Tucson, AZ 85719 USA

ABSTRACT

For the past eight years, the author has been examining trends in access to international scientific literature in major international bibliographic databases available on various information systems. A major portion of the research program examined and compared the languages of the documents and countries of publication for items published between 1970-1990 and recorded on MEDLINE, PsycInfo, BIOSIS, Chemical Abstracts, and several other databases on the DIALOG system. The second phase of this study is to examine the remaining scientific databases on DIALOG, including MATHFILE and AGRICOLA. A comparison of the international range of MEDLINE and EMBASE has also recently been completed. In order to attempt to assess actual amounts of literature produced, the Unesco statistics for literature production have been examined. In the course of this research the author encountered a variety of system attributes that affect the ways by which the Third World literature can be identified. Some of the policies and procedures that affect the inclusion of Third World science have been identified.

RESUME

Durant les huit dernières années, l'auteur a examiné les tendances de l'accès à l'information dans les principales bases de données de bibliographie scientifique internationale sur divers systèmes d'information. Une large partie de ce programme de recherche a consisté à examiner la langue des documents et les pays de publication et effectuer des comparaisons sur MEDLINE, PsycInfo, BIOSIS, Chemical Abstracts, ainsi que plusieurs autres bases de données accessibles sur le système DIALOG. La seconde phase de cette étude étudiera les bases de données restantes, notamment MATHFILE et AGRICOLA. Une comparaison de la couverture internationale de MEDLINE et EMBASE a aussi été completée. Afin d'estimer la production scientifique mondiale, les statistiques de l'Unesco sur la production littéraire sont également examinées. Les caractéristiques des systèmes d'information qui affectent l'identification de la production scientifique des pays du Tiers Monde sont étudiées, de même que les procédures et politiques qui affectent l'inclusion de la littérature en provenance des pays en développement.

INTRODUCTION

The role that bibliographic databases play in the international communication of scientific and technical information is critical. These databases represent what we know about research in medicine, psychology, engineering, education, and a host of other disciplines. As indexes to the world's literature on these subjects, they also offer a picture of research being conducted in the world's countries in these disciplines. They enable the assessment of developments in any given country with reasonable precision.

My own research over the last eight years has focused on the "internationality" of documents represented in major US scientific and technical databases. While there are many studies of the coverage of individual countries, this is an attempt to provide a comprehensive picture, from the United States perspective, of international developments in various scientific disciplines. International, for the purposes of this program, is defined as the languages in which the documents are written, and the countries from which they are published.

These are not the only measures of the "internationality" of science. Conference participation, funding patterns, and other forms of scientific communication were not examined. Publishing preferences in some disciplines for European and US journals are acknowledged but not addressed. The rapidity with which journals have changed journal titles from local languages to English, and increasing acceptance of papers only in English were not investigated. However, the assessment of languages being used, and sources of documents, adds to the current picture.

The overall objectives of the research program are (a) to determine the trends in the use of various languages being used to report scientific and technical findings in various scientific disciplines, (b) to determine the trends in the contributions by various countries in these disciplines, and (c) to determine trends in nationalities of authorship in these disciplines. The project has focused on the major bibliographic databases available on the DIALOG system, of DIALOG Information Services, Inc., a vendor of more than 350 different databases located in Palo Alto, California. This system was used because it is the US system with the broadest range of disciplines represented.

Necessarily, a fourth problem must be addressed, and that is the degree to which the databases do provide comprehensive international coverage of their disciplines. The databases cannot identify and include every single scrap of information in their disciplines. It would not be cost-effective to do so. The databases address the "important" literature for the US scientist. However, how much literature is being produced in medicine, or biology, or chemistry, is simply not known. The existence of the Hispanic American Periodical Index, and the Index Medico Latino-Americano suggest that there are literatures of interest to others in other countries.

In the course of the project, system and database attributes were identified that made it in some cases simple, and in other cases more difficult, to conduct the research itself: and these are the focus of this paper.

The research described below is a report of work in progress: the entire project will not be completed for another two years. Pieces are being reported as they are completed, however.

SUMMARY OF FINDINGS

The First Eight Databases

The detailed findings of the first part of this research project have been reported elsewhere[1] in general, the inclusion of materials in non-English languages has declined significantly over the last twenty years. While in 1970, in MEDLINE, for example, 60% of all records were for documents in English, in the late 1980's the proportion is closer to 80%. The proportion of records for English documents in Chemical Abstracts rose from 55% to about 65% in the same period. Similar increases were found for virtually all of the databases examined. Documents in French, Spanish, Russian, Portuguese are all declining in number. The two exceptions are Japanese and Chinese documents. In virtually all files, these two languages increased in presence, particularly after 1980. (A chart of individual languages studied and their overall trends is presented as Table 1 in the Appendix.)

Examining country of publication data illustrates the first of many problems identified in the course of research that inhibit the easy use of these databases to assess international contributions in various areas of research. DIALOG bibliographic records are organized into fields, such that the author is a separate piece of text from the title, for example. Not all databases contain the same fields: and the country of publication field is one that is not always present. As detailed below, this means that the information is either not readily accessible, or not available at all. Of the eight databases initially examined, only three contained the field. Table 2 of the Appendix indicates that, with the exception of materials from Japan and the People's Republic of China, materials from outside the United States generally have been in decline. Records for materials from the United States comprised 30% of the MEDLINE database in 1970; more recent figures are 45%.

The preliminary examination of eight scientific and technical databases focused on the United States and Europe because the study was targeted at the most frequently represented countries and languages. The data for countries making up 90-99% of a given database were examined. Varying percentages of

---

[1]Gretchen Whitney, Language Distribution in Databases: An analysis and Evaluation, Metuchen, NJ: Scarecrow Press, 1990.

the databases were examined because varying numbers of countries contributed 90-99% of the database. This meant that records for documents from only Japan, the People's Republic of China, and in one instance India were included beyond the European, North American, and Oceanic communities.

In the MEDLINE portion of the study, for example, the nine countries most frequently found in the file were studied. All of these countries contributed 2.5% or more of the total records to the databases. To bring the study up to 99%, an additional 31 countries would need to be examined. These countries would have included India, China, Brazil, South Africa, Israel, Mexico, Argentina, Chile, Turkey, Taiwan, Thailand, Singapore, and Egypt. They (and a few other European countries) contributed 9% of the database. The remaining 57 countries in the database contributed 1% to the total file. There is no reason why the methodology could not be applied to further countries: it simply has not been done yet.

The same approach (the most frequent contributors) was used to identify languages. BIOSIS, for example, actually contained 54 different languages. Twelve languages were studied, which comprised 99% of the database. There were generally over 10,000 records for documents in each of these languages. There were but 149 documents in Thai, for example, 45 in Hindi, and 233 in Indonesian.

India, Kenya, China and Brazil in MEDLINE

Due to the recognized concentration of effort on the European community, additional countries -in specific Third World countries, were investigated for this report. An examination of the data for documents from four Third World countries suggests that these may not mirror the declining patterns shown in many other instances, and records for documents from these countries require further research. As noted below, however, this can only be done for a few of the databases, because the information is simply not available in the records.

Brazil, India, and Kenya were the most frequently referenced countries of publication for their geographic regions. China was selected as the next-largest non-English speaking country in the region. Records for documents from Brazil form a roughly stable pattern, ranging from 800-1200 documents per year. Kenya and India have stronger showings in the 1980's than they did in the 1970's, as the following charts show.

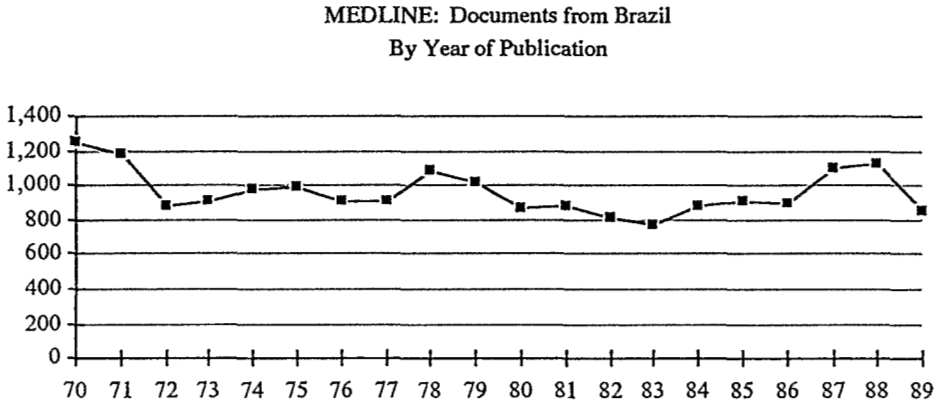

Chart 1: MEDLINE Documents from Brazil

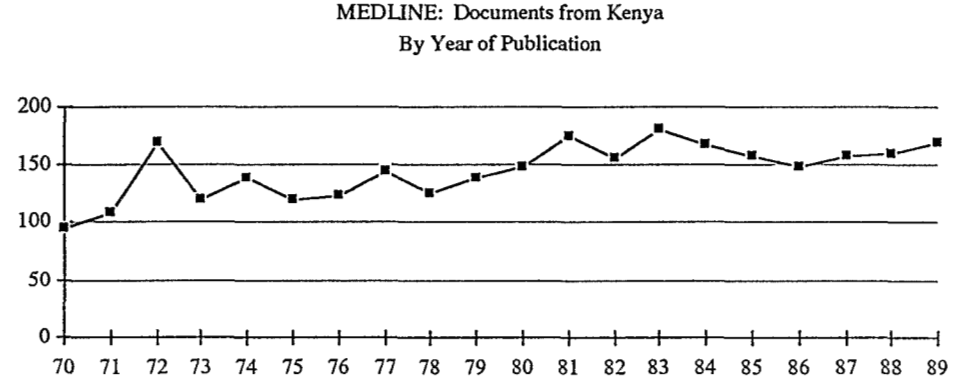

Chart 2: MEDLINE Documents from Korea

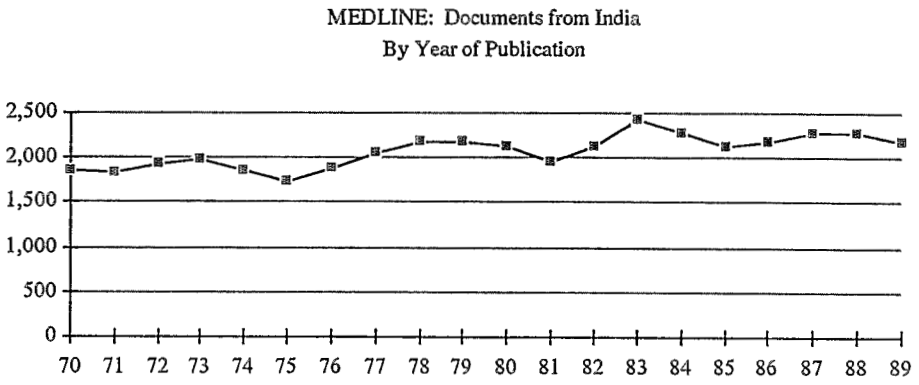

Chart 3: MEDLINE Documents from India

The chart for China shows a familiar pattern for documents from China in many databases: it is virtually absent until 1978, rises sharply in numbers until 1986, and stabilizes around 3500 documents per year after 1986.

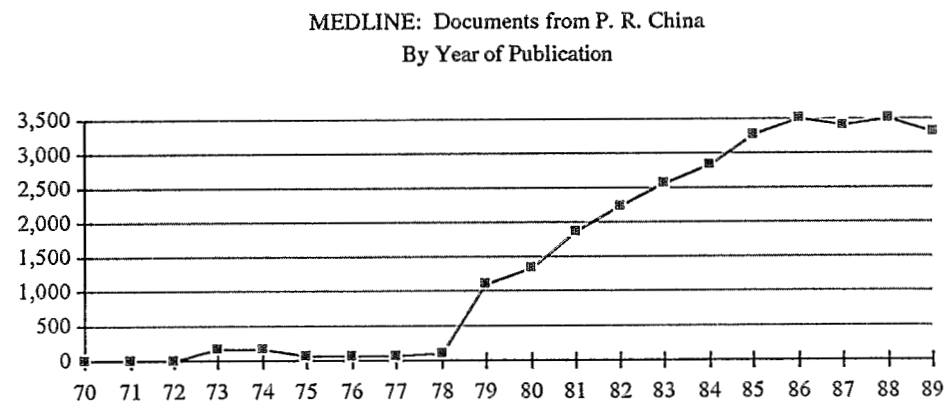

Chart 4: MEDLINE Documents from the People's Republic of China

Database Comparisons: MEDLINE and EMBASE

The structure of searching in DIALOG not only enables comparison of the quantity of research identified by different databases and different subjects, it also enables comparison of databases in the same subject. A recent project compared

MEDLINE and EMBASE, using the same language and country of publication fields.[2] EMBASE is generally considered the more "international" of the two. The results of the study revealed that in fact that was true (according to these specific criteria) up until 1984, when quantitatively MEDLINE became more so. That is, there were more records for documents published outside the US, and in languages other than English, in MEDLINE in more recent years. Details are presented in Table 4, in the Appendix.

Actual Literature Growth

The assessment of the actual growth of literatures in various disciplines is more problematic. Two resources were identified for study: the Unesco book production statistics, as reported in the annual Unesco Statistical Yearbooks, and the Online Union Catalog of the Online Computer Library Center in Columbus, Ohio USA. The latter project still is in the data gathering stage, and there is nothing to report at this time.

Examinations of the Unesco book production statistics[3] have suggested that while the data are incomplete, the trends are the opposite of what has been found in the databases. There is an increasing volume of literature in French, and Russian, and in Spanish for example. There are increases in disciplines such as medicine, the social sciences, and the sciences.

Asian production of pure science titles is increasing, as the chart on the following page shows.

Brazil, India, and Kenya were the most frequently referenced countries of publication for their geographic regions. China was selected as the next-largest non-English speaking country in the region. Records for documents from Brazil form a roughly stable pattern, ranging from 800-1200 documents per year. Kenya and India have stronger showings in the 1980's than they did in the 1970's, as the following charts show.

South American production is not. However, South American production gains show increases in the social sciences. Again, the chart is to be examined cautiously: the large number of countries reporting no data, or reporting erratically, affects the figures. A total of 80 countries reported this data worldwide in 1973, and maintained reporting in the late 1970's. Currently, only 60 countries report, and it is seldom the same 60 countries. If more countries reported, the increases would be even stronger.

---

[2]Gretchen Whitney, The internationality of Medical Literature: EMBASE and MEDLINE, Unpublished ms., 1990.

[3] Gretchen Whitney, "The Unesco Book Production Statistics", Book research Quaterly, Winter 1989-90, 5(4): 12-29; Gretchen Whitney, "The subjects of the World's Books: A further Examination of Unesco's Book Production Statistics", Book Research Quaterly, forthcoming.

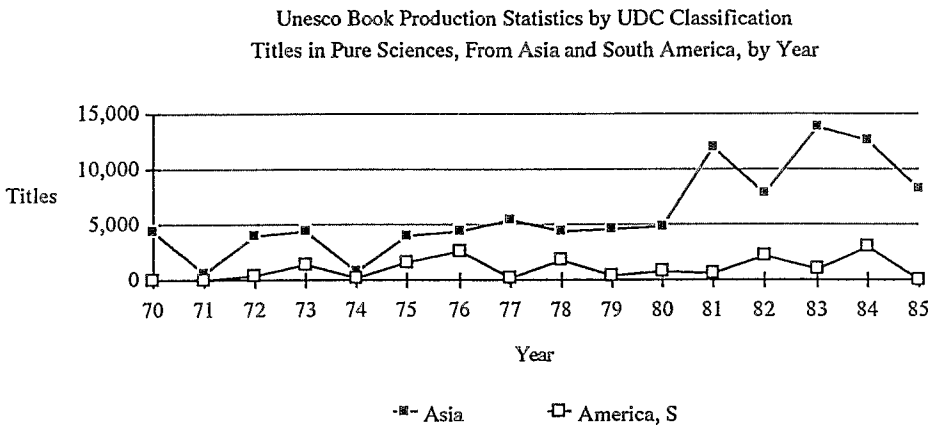

Chart 5: Unesco Book Production Statistics: Pure Science Titles from Asia and South America

A complementary project, investigating the Unesco statistics for serials, is currently underway.

In summary

Evidence gathered demonstrates that the inclusion of non-US publications, and those in non-English languages, has generally declined in major US scientific and technical bibliographic databases over the past twenty years. The exceptions are those materials from Japan and the People's Republic of China. Shifts in "internationality," as defined for the purpose of this project, are also occurring at the database level for biomedical information, in that MEDLINE is now quantitatively the more international file. Attempts to determine actual literature production have suggested, despite incomplete data, that in some cases the literature of given disciplines from different geographic regions is, in fact, increasing.

Feasibility of Using Bibliographic Databases for Research: System Software

The problem areas identified relate to (a) the DIALOG system searching software and that of the Online Computer Library Center, (b) the databases themselves, and (c) organizational variables such as collection development policies that affect the development of the files. The latter treats the Unesco book and serial production statistics as a data set.

DIALOG System Software

The DIALOG searching capabilities are strongly supportive of this type of research. Specific fields of information are readily searchable, multiple databases can be searched at the same time, and Boolean operators ("AND", "OR", and "NOT") are fully implemented to identify and clarify the use of values in different fields.

The only difficulty encountered is a problem with the algorithm that is used to identify duplicate records. Minor variations in titles cause the system to evaluate two records as different, when in fact they refer to the same document. This inhibited progress on an overlap study of two biomedical databases, MEDLINE and EMBASE. The problem was reported to DIALOG.

OCLC System Software

The ready ability to conduct this type of research on DIALOG is contrasted with the difficulties inherent in a similar project on the OCLC Online Union Catalog.[4] The Catalog, a database of 22 million bibliographic records representing the holdings of 10,000 libraries in the United States and abroad, allows for searching by the title, author, and several numeric fields only. This was by design. It was established as, in part, a cataloging system, and searching was to be "known item" searching for a work already in hand. Additional features such as subject access were planned for a later date. Languages and countries of publication also are inaccessible. My examination of serials records, therefore, necessitates the assignment of a programmer to write software specifically to extract the needed records.

That the new (as of January 1990) OCLC service, EPIC, is a full copy of the Online Union Catalog (without holdings), with DIALOG-type search capabilities. Language and country of publication are separate fields that could be searched readily to assess the holdings of Third World publications in libraries in the United States and abroad.

Feasibility of Using Bibliographic Databases for Research: Database Characteristics

While vendors, such as DIALOG, OCLC, and Unesco, offer databases and provide access to them, the databases are constructed by third parties. These database providers also affect their usefulness for research, because they are the ones that identify the fields for each record, and provide the data for each field.

---

[4] Gretchen Whitney, "The Language Distribution of Serial Records in the OCLC Online Union Catalog", OCLC Library and Information Science Research Grant Program Award, 1989-1990.

They also develop policies which affect the inclusion or exclusion of the materials for which records are developed.

ISI's Science Citation Index and Social Science Citation Index

While ISI has been used heavily as a resource for the analysis of the growth or incidence of country's literatures, its problems also have been documented. This work complements that of other researchers studying the ISI data, by addressing a similar set of problems from a different direction: that of the bibliographic databases constructed not from attributes of the literature itself, but from databases built by national libraries (MEDLINE, from the National Library of Medicine), by professional associations (PsycInfo, from the American Psychological Association) and other third parties. This "participation" by the third parties does introduce additional variables affecting the picture of given disciplines that these databases offer, and that is discussed below.

Selectivity in Databases

Necessarily, secondary services must be selective in the titles that they choose to index. It costs money to hire the indexers, and then to manage and print the results of the indexing and abstracting effort. The secondary services are under various pressures to curtail production costs. Sometimes, however, this selectivity affects the international character of the database. For example, Aspen Systems, Inc., the creator of the National Criminal Justice Periodicals Index, was directed by its monitors in the US National Institute of Justice in the middle 1970's to cut the number of records entered into the file, for several successive years. The foreign materials were the most expensive, difficult to acquire and to process; so acquisition of these materials was virtually ceased. This resulted in the following pattern of French-language materials:

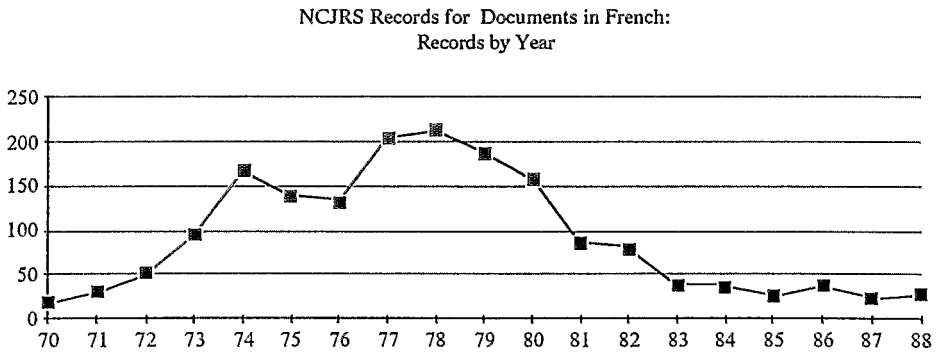

Chart 6: NCJRS Records for Documents in French

The staff, when interviewed in the mid 1980's, expressed great concern over these materials. Their numbers have not increased, however. There were only 47 items added with a publication year of 1988, as the chart indicates: and only seven with a publication year of 1989.

A similar impact was felt on materials from outside of the United States, and India will serve as our example:

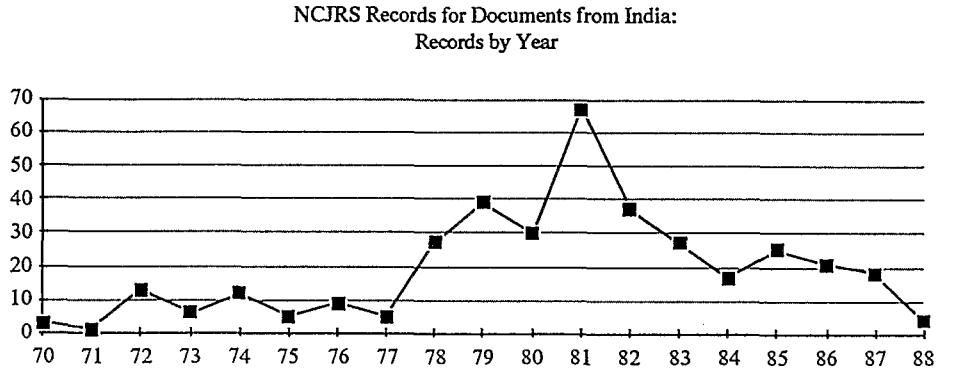

Chart 7: NCJRS Records for Documents from India

Eight materials were added with a publication year of 1988, and none with a year of 1989 or 90. (The database appears to have a significant lag-time in processing materials. As of this writing (August 1990) there are no materials in the database with a publication year of 1990.)

Fields and Field Tags

DIALOG made a substantial step forward in 1985 when it implemented field tags for display of records. While you could search for specific information in specific fields before 1985, the printed record, upon retrieval, had no tags. This made it quite impossible to download or capture records to a disk file for later processing in, for example, microcomputer software. With the implementation of DIALOG2, such processing and analysis is easily undertaken.

DIALOG in the last five years also added formatting capabilities, including a capability for specifying individual fields for output. For example, a set of records representing research produced from Thailand could be identified, and their full CS fields alone could be output for analysis.

Particularly relevant fields to this research are the language of the document, the country of publication, and the author affiliation or corporate source. The utility of the language field is readily apparent. And, straightforward analysis of countries of publication, or source countries, is also quite simple. All of the

values in a given field can be examined by using an "expand" command. These identified values then can be paired with, for example, year of publication, to assess trends.

Combining these fields (language and country of publication, or country of publication and source country for author) would yield interesting results indeed. A comprehensive approach would be somewhat more tedious, because all of the countries in one field will have to be paired with all of the countries in the second field. It is not possible to identify a set of records (for example, those with Brazil in the country of publication field) and then examine the values in the language field for just those records, without printing them all out. (Technically, that is possible: but it would yield unwieldy numbers of records.) A sampling strategy is being developed to work on this problem.

Finally, not all databases contain the relevant fields identified above. While all do include a field for language and some sort of corporate source data, only Aerospace, Agricola, GeoRef, MEDLINE, NCJRS, NTIS, PAIS, and PsycInfo include fields that can be examined for corporate source or author affiliation data. And these are inconsistently defined across databases.

Values in Fields: Multiple Forms of Languages, Country Names

In most of the files, the language field is "clean," in that it contains only the name of the language. In some files, however, the field includes extraneous codes and data which must be removed for analysis. The same can be said of the country of publication field. The corporate source field, however, usually contains the entire source name (such as "Dep. Physiol., Fac. Sci., Mahidol Univ., Rama VI Road, Bangkok 10400, Thailand."). In determining the countries or institutions represented in the file, the EXPAND command must be used to list all values represented, and the individual countries must be identified. To identify specific countries, it is a simple matter to search for individual country names. MEDLINE encodes country names, and these must be interpreted from written documentation. (Text-searchable words are available from 1966-1974 and after February 1987.)

Not all databases are consistent in the designations that they use for names of languages or names of countries. In one database (Inspec), for example, there were multiple "versions" of various forms of language names ("Spanish", "Spansih", "Spansish", "Spnaish"; the same database contained eleven forms of "Russian"), and different spellings of country names. While these errors generally do not affect large numbers of records (in the Spanish example, 5 records were incorrect, 2,889 were correct), they do represent a problem to be taken into account.

There may be extraneous or coded information in the field, which can affect the number of records retrieved. In one of the databases (Agricola), the total number of records could differ by several thousand. The database provider

should be contacted to determine, for example, if the coded records should be ignored in the total, or if they should be counted as English.

Databases have not always coded materials for language: one file (BIOSIS) did not start coding all materials until the middle 1980's. It is necessary, therefore, to calculate the percentage of records that have been coded with a language or country value.

To determine the percentage of non-English materials in a database, it is necessary first to determine the extent of English materials. Some databases explicitly note "English" for English language documents, others do not do so if the document is in English. And some databases have changed policies over time, which offers further challenges.

MEDLINE serves as an example of a database that uses a highly structured, hierarchical system of country codes for its country of publication field, and this structure must be understood and used in any analysis. The incidence of duplicate country codes must be managed as well.

Separate Vendor Handling of English and Non-English materials

There is an increasing trend to separate out English-language materials from non-English materials, to assist libraries and users in purchasing only the data that they really need. The Online Computer Library Center, one of the largest bibliographic cooperatives in the United States, for example, offers a CD-ROM-based cataloging service. Non-English materials are on a separate disk, and are purchased separately. In the middle 1980's, the American Psychological Association, creator of PsycInfo, one of the major secondary services, decided that the printed version would contain only the English-language materials, the foreign-language materials would be available only in the electronic version of the database. The measure was taken to save production costs for the printed service. While this does mean that the non-US materials are indeed electronically accessible to the rest of the world, countries may find that the printed versions are less expensive to acquire and use.

The OCLC Online Union Catalog

As a cooperative venture among thousands of libraries, the Online Union Catalog has been in development since it first went online in 1971. With 22 million records, it is perhaps the largest bibliographic file in the country. With so many participants in its development, and its immense size, errors and duplicates are unavoidable. OCLC has, at least since the middle 1970's, maintained a program of collecting error reports from participating libraries and diligently working on them. Detection of duplicates is a continuing item on the research agenda, and researchers currently are working on various algorithms for standardizing and improving records for serials. The current research project on the linguistic distribution of serials in the Online Union Catalog will, in part,

further support the effort to determine the broad numbers of serials being published in various languages.

OCLC uses derived search keys (pieces of text from the target field, such as the first four characters of the author's last name as a part of the author/title search key) for the Online Union Catalog, and the language field is not one of those fields from which a search key has been taken. Therefore, direct searching for records in specified languages was not possible. Special programs are being written at OCLC to pull out a sample of the database for study. The small sample will be processed using standard microcomputer software.

The recently-released EPIC service does permit searching by language and other relevant fields. It was not available at the time this particular proposal was submitted and awarded.

The Unesco Statistics

The Unesco book and serial production statistics are important as indicators of not only scientific and technical developments in various countries, but they reflect the humanities as well. As such, they have the potential for enabling us to assess international work in all of these areas, and the progress being made to support educational, scientific, and literacy efforts around the world. Unfortunately, many countries, for reasons of lack of available data, or a lack of an internal reporting system, or political reasons, do not participate in Unesco's program to collect cultural data. My own country is no exception. But assessments from this potentially valuable resource will remain uncertain without wider participation in collection of the data.

Feasibility of Using Bibliographic Databases for Research: Organizational Variables Affecting Database Development

Organizational variables include the establishment of or changes in policies regarding what materials will be included in a database, corrections made to the file, and changes in the database developer's relationship with other agencies. These are frequently the most difficult to identify and relate to the database, because they are often buried in the archival memory of the agency.

The participation of the national libraries of England and France in the Online Union Catalog clearly will increase the visibility of their records in the database, and increase their accessibility for United States libraries. The national libraries will benefit mutually.

When Questel, the French database vendor, opened offices and a local telecommunications node in Virginia in the United States, US researchers had increased access to French databases.

Other examples include the separation of English-language and non-English records in PsycInfo, and the policy changes mandated by the National Institute of Justice, noted above. Finally, the incompleteness of the Unesco statistics prompted contact with Unesco itself, to determine why the data were so incomplete. While this does not make the data any more complete, it helps us to understand where the problem lies.

An analysis of the Population Bibliography, produced by the Carolina Population Center at the University of North Carolina, USA, provides a dramatic example of how difficult it can be to identify these variables.[5] A preliminary analysis of the language and country trends in the database could not readily be explained by funding patterns in the field, international contacts and agreements by the Center itself with centers in other countries, or acquisitions policies. Only a few years could be explained by staffing patterns, as determined by examination on site of the archives of the library. Only upon review of the draft of the paper resulting from the research did the librarian remember that, in preparation for sending the file to DIALOG, large numbers (50% in some cases) of non-English and foreign-produced records were removed from the file. The explanation for the patterns seen in the database were not in public or archival records: they existed only in the memory of the librarian. If the patterns as seen in the database were taken at face value, and used to understand the growth of the literature of population and family planning, they would have been quite wrong.

In this example, the author had the favored advantage of having worked at the Center during the years that the database was being developed, and staff critical to the study were still in place. If someone else had undertaken the project, the results could have been quite different.

---

[5] Gretchen Whitney " Organizational Variables Affecting the Conduct of Bibliographic Database Research: Factors Influencing the Internationl Coverage of databases", Information Processing and Management, forthcoming.

This is by no means to suggest that all database studies should only be undertaken by former employees of the developing organization. Rather, it is to re-state that databases are products of human beings, and organizations, and that it is important to look beyond the raw facts whenever possible.

CONCLUSIONS

Bibliographic databases such as those on DIALOG, and the Online Union Catalog and EPIC databases, are well suited for researching the output of Third World countries. The fields are available for examination, the searching capabilities are present. In some cases, custom software must be developed. There is a need to be concerned for, and account for, inconsistencies in data entry in some cases, the use of special codes, and the general development policies under which the database has been developed. Finally, it is important to take into account, as much as possible, the organizational variables which affect the database and its continuing development.

Aknowledgements

The research describing DIALOG system capabilities was and continues to be funded in part by a research grant from DIALOG Information Services, the focus of which was the determination of linguistic and source trends in scientific and technical bibliographic databases. The author is most grateful for this support.

Table 1: Trends Individual Languages Studied in the First Phase of the Project, by Database (Key to Abbreviations in Table 3)

	Database							
Lang.	Bio.	CA	Geo	Med	NCJ	OCE	PAIS	Psyc
Chinese	+	+	+	+		+		
Japanese	+	+	+	+	-	E		+
Bulgarian	+	-		+				+
English	+	+	+	+	-	-	-	+
Spanish	-	+	-	+	-	E	+	+
Korean	+					E		
Slovak		+		E				
Dutch				+	-			+
Danish				+				
German	+	+	-	-	-	E	-	+
Norwegian				-				
Romanian		E	-					
Portuguese	-	E		E		E	-	+
Serbo-Croatian				-			E	
Ukrainian		-						
Czech		+	-	-				E
Hungarian		E		-				-
Russian	+	-	-	E		+		E
Polish	+	E		-	-	E		E
Italian	+	-		E	-	E	-	+
Hungarian		E		-				-
Swedish				-	-			
French	+	-	-	-	-	+	-	+

+ = increased, - = decreased, E = too erratic or stable to classify

Table 2: Trends Individual Countries Studied in the First Phase of the Project, by Database (Key to Abbreviations in Table 3)

	Database							
Country	Bio.	CA	Geo	Med	NCJ	OCE	PAIS	Psyc
U. S.			+	+	-			
Netherlands			E	+	-			
U. K.			_	+	-			
China			+					
Japan			+					
Italy				+	-			
Australia					-			
Belgium					-			
Canada					-			
India					-			
Sweden					-			
Poland				-	-			
Switzerland				-	-			
France			E	E	-			
W. Germany			E	-	-			
Russia			-	-	-			

+ = increased, - = decreased, E = too erratic or stable to classify

Table 3: Abbreviations from Tables 1 and 2

Bio.	BIOSIS
CA	Chemical Abstracts
Geo	GeoRef
Med	MEDLINE
NCJ Service	NCJRS (National Criminal Justice Reference
OCE	Oceanic Abstracts
PAIS	Public Affairs Information Service
Psyc	PsycInfo

Table 4: Quantitative Dominance of EMBASE and MEDLINE Languages and Countries of Publication, 1974-84 and 1985-89

Language	EMBASE 1974-84	EMBASE 1985-89	MEDLINE 1974-84	MEDLINE 1985-89
English			*	*
German	*			*
French	*			*
Japanese			*	*
Russian			*	*
Italian	*			*
Spanish			*	*
Polish			*	*
Chinese			*	*
Country of Publication				
US			*	*
USSR			*	*
W. Germany	*	*(7)	*	
England			*	*
Japan			*(8)	*
France	*			*(9)
China			*	*

The asterisk indicates which database contained the largest number of records in the given language or with the given country of publication during that time period.
(7) computed from the spreadsheet, 105K to 102K records.
(8) computed from the spreadsheet, 123K to 127K records.
(9) computed from the spreadsheet, 43K to 46K records.