Analyzing the Journal Coverage of Abstracting/Indexing Databases at Variable Aggregate and Analytic Levels

Péter Jacsó
University of Hawaii

Several studies have attempted to reveal the real scope, completeness, and consistency of journal coverage across time in abstracting/indexing publications. Online and CD-ROM databases make easier the collection of posting information for journals to determine the traits of their coverage but do not eliminate entirely the tedium of manual methods. Difficulties are compounded when multiple databases are used to compare the journal coverage and overlap of several databases for an extended time period. This article discusses the use of advanced search commands and the Journal Name Finder database of DIALOG, to simplify the collection and processing of posting information for 42 prestigious library and information science (LIS) serials between 1966 and 1996 in six databases. It presents the major types of deficiencies in journal coverage that may yield very incomplete search results and may distort the results of bibliometric and scientometric studies. The methodology allows the creation of various subsets by the versatile combination of journal titles, time periods and databases, and encourages longitudinal data collection and comparison of variously aggregated datasets. It can be used for profiling the journal coverage in the evaluation of databases in support of bibliometric research, database acquisition, and licensing decisions.

Online and CD-ROM abstracting/indexing and full-text databases have become the mainstream tools in libraries to find information about publications by a variety of criteria and their combinations (subject, author, author affiliation, language, type, country or year of publication, etc.). The number of journals and other periodical publications covered by a database—ranging from a few hundred to several

thousands—is often indicated in the introductory screen or in the help file of a database. However, there is usually no indication for the extent of coverage of the journals that would inform the users about the time span and depth of their coverage (Jacsó, 1992). Printed guides and journal lists may include the first and last year of coverage of individual journals. The identification of core journal status may suggest that the source has complete coverage. This information, however, is not readily available for, and/or is not studied by, the casual users (Omaji, 1994) who may assume that what they found in a database represents the published literature. Professional database users and experienced researchers are more skeptical in their assumptions, and are in a better position to get hold of and study user guides and journal lists. However, the information presented by database producers and publishers are not detailed and informative enough to learn the true dimensions of database coverage (Jacsó, 1994).

Knowing the scope, comprehensiveness, consistency, and currency of journal coverage in databases is of special importance for bibliometric and scientometric research projects that use electronic databases to study publishing patterns and citation behavior, to rank authors, institutions, and journals, to explore research fronts and trends in a discipline where journal publications are the predominant information sources (Duff, 1995; Kabir, 1994; Pettygrew & Nicholls, 1994; White, 1990; Yerkey, 1993). Heavy fluctuation and significant gaps in coverage, shallow, tardy, and prematurely terminated coverage of purportedly core journals may distort the results of bibliometric studies, and yield insufficient results for all types of literature searches. Studies about the journal coverage of library and information science (LIS) databases are of obvious interest for the LIS profession, and may serve also as a model for similar studies in other disciplines that have borrowed ideas and research methodologies from LIS (Cronin & Pearson, 1991).

Many studies have analyzed the coverage of LIS journals even before the availability of online databases. The ones by Gilchrist (1966), Goldstein (1973), and Edwards (1975) stand out from that period for their comprehensiveness. Typically, however, the testing of the coverage was limited to journal title level, that is, to the determination if a journal is covered or not by a service, and did not discuss the extent of coverage of the journals. Even those studies that used a set of journal articles from comprehensive bibliographies as a benchmark were limited to relatively small sets due to the tedium of the process (La Borie & Halperin, 1981). A notable exception for large-scale analytic-level evaluation was the study by Flanagan, Kennedy, and Wood (1973), a team of researchers from three of the largest abstracting/indexing services with access to the data on their mainframe computers. They compared the article level coverage of BIOSIS, Chemical Abstracts, and Compendex.

Goldstein's (1973) study was replicated and significantly enhanced later by La Borie, Halperin, and White (1985) who examined the trend in coverage of the same four abstracting services, along with six additional services from the field of business, engineering, social sciences, and humanities. They also created an innovative map for visualizing the overlap among the 10 services.

The availability of computer readable versions of abstracting and indexing publications from personal computers made the testing and analysis of journal coverage somewhat less tedious, and article level analysis such as that of Poyer (1982) started to appear. The analyses focused on comparing global intake of articles from journals; they used relatively small samples in terms of time period, number of abstracting/ indexing services, and/or journals analyzed. Other studies used several online sources to find the most appropriate databases and the most productive journals for a specific subject, such as CD-ROM applications (Nicholls, 1989; Schwartz, 1992). These studies provided useful indication for the extent of coverage of several journals, but were not meant to reveal the pattern of coverage throughout an extended period of abstracting and indexing the primary journals by the database(s). Not even the CD-ROM versions of the databases proved to be a perfect solution for journal coverage analysis. Although they provide the convenience of unlimited usage their software did not offer such powerful commands as those of the online DIALOG, or the now defunct ORBIT systems. Jacsó (1992) tested the coverage of several core journals in three general interest periodical CD-ROM databases using a rather time consuming method of data collection to illustrate the wide differences in the extent of coverage of the sample titles across the entire time-span of the databases.

Though online and CD-ROM versions of abstracting/indexing and full-text databases opened new horizons for analytic-level journal coverage studies, some stumbling blocks remained for simplifying data collection and analysis for in-depth journal coverage studies. Perhaps the inconsistencies of journal name spelling remained the most significant barrier. These make the entries widely scatter in the phrase indexed journal name field of a database. Long journal names are particularly troublesome because they are not only often truncated, but also trigger non-standard and inconsistent abbreviations. These are just some of the examples of the more than 30 variants found in a dozen of DIALOG databases for the *Journal of the American Society for Information Science*:

- Am Soc Info Sci J
- Am Soc Info Science J
- American Society for Information Science Jour
- J. Am. Soc. Inf. Sci.
- J Amer Soc Info Sci
- J Amer Soc Inform Sci
- Journal American Society for Information Scien
- Journal of American Society for Information Sc
- J. of the Am. Soc. for Information Sci

Even with defensive searching that contemplates journal name variants, it is easy to miss a highly posted variant and it can grossly distort the results. Some inconsistencies are predictable, but many are not, and in databases with thousands of

journal name entries, browsing through the index to find all the variations of the name of the journals keeps the process tedious, error-prone, and expensive.

This article describes new methodologies and techniques that alleviate most of the problems encountered in evaluating and comparing journal coverage of databases. Additionally, it presents some of the findings of a research project using *Information Science Abstracts (ISA)* as the target database. The methodologies and techniques to collect posting information for journals and other serials included the use of the special databases and advanced commands of DIALOG, such as Journal Name Finder, DIALINDEX, OneSearch, Rank, IDO (Identify Duplicates Only), and RD (Remove Duplicates). Word processing and spreadsheet programs were used to generate search scripts for unattended and consistent execution of time-consuming queries, and to process and present the data in a flexible and visually informative manner at various levels of aggregation by journal categories, database groups, and time periods. The methodology presented may serve as a model for benchmarking the journal coverage of bibliographic and full-text databases of any journal subset in any discipline.

METHOD

The methodology described here puts the journal coverage of the target database *(ISA)* in proper context by mapping it across a variety of discretionally chosen subsets (e.g., by journal categories or country of publication), various time spans (past year, past 5 years, or entire database time-span). For certain criteria (such as extent of coverage) the journal coverage of a database can be best described and evaluated in relation to its peer databases.

DIALOG has been used for this research for three reasons. First, it has incomparably powerful software features and special databases for benchmarking purposes. Second, it was the only online service that provided access to *ISA* at the start of the research project (now SilverPlatter ERL service also offers *ISA* online). Third, Dialog has the most databases that can qualify as members of the peer group of *ISA*. The only highly relevant database not available on DIALOG at the time of the research was *Library Literature (LibLit)* of the H.W. Wilson Company. By the summer of 1997 *LibLit* also became available on DIALOG, and reconfirmed the validity of the data provided by H.W. Wilson earlier from their in-house database.

The results presented here for all the databases—except for *LibLit*—reflect their update status as of January 1, 1997 providing a natural deadline and level playing field to compare coverage up to and including 1996. *LibLit* data were available up to the November, 1996 update.

Selection of the Journals

The most "current" bluesheet of *ISA* available on DIALOG's Web site as of January, 1997 was last updated in July 1994. It claims that *ISA*—in addition to other

documents—"indexes and abstracts articles from over 300 journals." A more current, 1996 pamphlet by IFI/Plenum, the datafile producer, claims that "*ISA* scans more than 400 international journals."

Table 1 lists the 42 serial publications selected for testing twenty-one of them are identified by *ISA* as core journals that are "abstracted in their entirety." These are the category 'A' titles. The 21 other items in the sample were selected to include *ISA* non-core titles. These are prestigious and high impact journals of library and information science. Thirteen of these 21 titles were among those top ranked by deans of 47 library schools accredited by the American Library Association (ALA), and directors of 43 libraries of the Association of Research Libraries (ARL) in the classic study of Kohl and Davis (1985) and in the Tjoumas and Blake (1992) survey of the perception of professional journals by library school faculty specializing in school and public library topics. These are identified as category 'B' titles. The remaining 8 non-core *ISA* titles (category 'C' titles) represent prestigious non-U.S. journals and one important annual proceeding. (The *Annual Review of Information Science and Technology* and the *Proceedings* of the Annual Meeting of the American Society for Information Science are technically not journals but are listed as such in several databases, and are included by the term "journals" in this article.)

While the 23,835 records retrieved from *ISA* by the sample titles represent only 12.4% of the entire population (191,474 records) of *ISA*, the subset of the sample set of records without items for the two annual publications mentioned above, constitute more than one–fifth of the 109,172 records identified as journal articles in *ISA*. Determining the total number of unique records in the entire *ISA* database is a difficult task as an unprecedentedly high portion of *ISA* consists of real duplicate and triplicate records. (As to be explained later not all records identified by DIALOG are real duplicates). Based on various samples, it is estimated that about 6% of the records are real duplicates or triplicates in *ISA*. No real duplicates were found in the same sample for the other databases in the peer group.

Selection of the Peer Group of Databases

Some aspects of the journal coverage of the target database (such as the depth of coverage) can be best assessed in the context of the coverage offered by other databases that abstract and index the same journals. These are the peer databases of the target database. The peer group of databases may be established based on previous research experience and recommendations in the professional literature. Traditionally, one or more of the following abstracting/indexing publications were used in comparisons with *ISA*: *Library and Information Science Abstracts* (*LISA*), *ERIC*, H.W. Wilson's *Library Literature* *[LibLit]*, and *Referativnuj Zhurnal*. Alternatively, the researcher may use and/ or customize the various predefined groups and supergroups of databases categorized by disciplines and subdisciplines by DIALOG. While for example, for

TABLE 1
Sample Journals with Postings for Databases in the Peer Group

Abbreviated Journal Titles	INSPEC	SSCI	LISA	Pascal	LibLit	ISA
ANNU REV INFORM SCI		284	44	196		252
ASLIB PROC	654	1,282	938	1,005	493	698
B MED LIBR ASSOC	355	2,126	973	887	922	1,171
CAN J INFORM LIB SCI	16	94	70		82	32
CAN LIBR J	111	2,446	602	503	1,004	475
COLL RES LIBR	24	2,753	871	696	961	985
DATABASE	724	1,993	968	556	1,158	840
EDUC INFORM	51	622	280	77	411	92
ELECTRON LIBR	462	909	467	376	608	429
GOV INFORM Q	155	833	316		444	262
GOV PUBL REV	52	1,594	442	571	222	
IFLA J-INT FED LIBR		928	532	377	724	189
INFORM PROCESS MANAG	863	1,748	924	802	1,042	987
INFORM TECHNOL LIBR	400	1,022	527	431	809	467
INT FORUM INFORM DOC	192	550	413	424	284	184
INT J INFORM MANAGE	264	626	280	204		162
INTERLEND DOC SUPPLY	95	364	267	260	298	73
J ACAD LIBR		2,324	972	3	1,286	516
J AM SOC INFORM SCI	1,208	2,092	1,217	1,107	904	1,544
J DOC	260	1,737	485	366	793	903
J EDUC LIBR INF SCI	60	539	294	265	517	220
J INFORM SCI	503	1,007	685	677	536	617
LAW LIBR J	25	1,580	479		511	63
LIBR ACQUIS PRACT TH	112	1,235	773	553	914	459
LIBR INFORM SCI RES	51	530	301	216	337	302
LIBR J		27,108	1,407	1,383	7,346	1,015
LIBR QUART	26	2,487	412	265	934	397
LIBR RESOUR TECH SER	243	1,327	471	385	830	793
LIBR TRENDS	175	1,066	835	872	540	907
LIBRI	28	504	460	496	289	352
NACHR DOK	566	1,636	521	927	341	500
NAUCH-TEKHN INFORM 1	902	2,315	245		1,109	1,032
ONLINE	1,240	2,207	1,152	777	1,232	998
P ASIS ANNU MEET	36	2,342	106	120		1,527
PROGRAM-AUTOM LIBR	559	1,004	529	500	599	176
RQ	48	4,005	685		1,284	604
SCHOLARLY PUBL		743	210		330	159
SCIENTOMETRICS	45	1,124	689	689		644
SERIALS LIBR	374	1,159	827	376	730	691
SPEC LIBR		1,657	801	842	576	1,788
WILSON LIBR BULL		7,717	1,136		3,973	409
Z BIBL BIBL		1,270	210	626		23
Total for 42 sample titles	10,879	90,889	24,636	18,665	35,722	24,159
Total without LJ and WLB	10,879	56,064	22,093	17,282	24,403	22,735
Category A titles (bold)	7,332	30,892	13,479	10,095	13,984	14,421
Category B titles (italics)	1,197	49,773	8,656	5,713	18,777	6,064
Category C titles	2,350	10,224	2,501	2,857	2,961	3,674

pharmacology there are more than 50 databases on Dialog, and a great variety of categories (Pharmacology [PHARM], Pharmaceutical Industry [PHARMIND], Pharmaceutical and Healthcare Industry [PHIND], there is only one category for Library and Information Science [INFOSCI] databases. It offers relatively few databases (8), and at least one of them (*British Education Index*) is practically irrelevant for LIS. At the same time, some highly pertinent DIALOG databases are not included in the INFOSCI category (*Social SciSearch, Pascal*, and *MathSci* to name a few).

For this research, five primary peer databases were selected. Additionally, data were collected from four other databases that have good coverage for some of the journals in the sample, but are not included in this discussion. The members of the peer group are listed in Table 2 along with additional data that will be referred to later in the text. It is to be borne in mind in evaluating the results that none of the other databases go back as far as ISA in their coverage.

TABLE 2
Peer Database Group Summary Data

	INSPEC	SSCI	LISA	Pascal	LibLit	ISA
Producer	IEEE	ISI	Bowker	CNRS	H. W. Wilson	IFI/Plenum
Connect time Fee	$60	$75	$30	$45	$40	$75
Type, print full record	$1.85	$1.95	$1.05	$1.40	$0.10	$1.15
Retrospectively (in years)	28	25	28	24	12	31
Total number of records	5,457,178	2,974,958	170,513	11,300,204	162,000	191,424
LIS subset	57,188	113,275	170,513	N/A	162,000	191,424
Sample journals covered	34	42	42	35	37	42
All 42 Sample Titles						
Past 5 Years	2,645	36,480	7,868	4,831	12,475	4.947
Share of Cummulative Total	3.82%	52.68%	11.36%	6.98%	18.02%	7.14%
Past Year (1996)	448	8,305	1,951	652	1,058	622
Share of Cummulative Total	3.44%	63.71%	14.97%	5.00%	8.12%	4.77%
Only Category A Titles						
Past 5 Years	1,950	8,902	4,915	2,928	5,440	3,858
Share of Cummulative Total	6.97%	31.80%	17.56%	10.46%	19.43%	13.78%
Past Year (1996)	369	1,989	636	473	1,257	489
Share of Cummulative Total	1.08%	38.15%	12.20%	9.07%	24.11%	9.38%
Only Category B and C Titles						
Past 5 Years	695	27,578	2,953	1,903	7,035	1,089
Share of Cummulative Total	1.68%	66.85%	7.16%	4.61%	17.05%	2.64%
Past Year (1996)	79	6,316	206	179	1,350	133
Share of Cummulative Total	0.96%	76.44%	2.49%	2.17%	16.34%	1.61%

Collecting Posting Information for Journals

The bane of journal coverage studies is the wide intra-database and inter-database scattering of names and abbreviations for the same journals. In this peer group only *Social SciSearch* (SSCI), INSPEC, *LibLit*, and *LISA* have appropriate authority files and authority control for journal names. *LibLit* was not available on DIALOG at the time of the research but H.W. Wilson made available the yearly postings information (up to November, 1996) for the titles in the sample in a format that could be imported in the three dimensional spreadsheet.

The Journal Name Finder (JNF) database offers a unique feature that ensures that even the most unpredictable variations of journal names can be found. JNF creates both word indexes and phrase indexes (with and without a prefix for the latter) from the *JN* and *JO* fields of bibliographic databases. This made it possible to formulate defensive search strategies even for the most inconsistently entered journal names, and also for the majority of the incorrectly spelled and abbreviated journal names. The word indexed entries in JNF allowed such defensive queries as *SELECT (J OR JNL OR JN OR JRNL OR JOURNAL) AND AM? AND SOC? AND INF? AND SC?* to retrieve all possible variations for the *Journal of the American Society for Information Sciences*. For some terms that would generate a very large set of records, it is wiser to spell out variations than to truncate after the first or second character. The searches in JNF were limited to the databases analyzed.

The results of JNF searches for all the titles in the sample were directly exported to a temporary spreadsheet. The searches in JNF used intentionally loose queries to allow the retrieval of all the variants of the sample titles. This yielded, of course, a number of irrelevant journals as well, but these were easy to remove from the list. A unique journal code was assigned to each relevant journal irrespective of its variations found in the index. Records were sorted by database identifier number as a primary sort key, and journal code as a secondary sort key. From the spreadsheet file a series of search scripts were generated for each title variations in each database. These scripts were then run against the appropriate databases. The results of the searches were cleaned, converted into an import format, and uploaded into a three-dimensional spreadsheet notebook. Each database had its own page as the Z-axis in the three-dimensional notebook. In each spreadsheet page the journal names were represented on the Y-axis, and the yearly postings on the X-axis (Figure 1). The spreadsheet notebook was used to provide the statistics and to produce the graphs for individual titles, a group of titles or for the entire sample and their various combinations by time period and databases from the peer group. This approach made it possible to create comparison tables and graphs in dynamic combinations of databases, journal titles and years that provide at-a-glance pictures of the absolute and the relative retrospectivity, duration, consistency and extent of coverage for a title or group of titles by the target database and by one or more of its peers. (Color coding of the series representing each database make the

FIGURE 1
Excerpt from the Spreadsheet Notebook

JC 202 ISA - 42 LISR titles	82	83	84	85	86	87	88	89	90	91	92
2 ANNU REV INFORM SCI		10	9	9	9	10		8	8	9	8
3 ASLIB PROC	18	61	25			3			67	76	81
4 B MED LIBR ASSOC			34	43	34	49	27	48	58	36	44
5 CAN J INFORM LIB SCI											
7 CAN LIBR J	39	50	29	22	34	33	34	35	17	40	11
8 COLL RES LIBR	26	33	34	20	27	39	20	76	29	30	16
9 DATABASE	23	31	43	33	37	44	56	22	32	46	67
10 EDUC INFORM		4	7	5	14	18	18	8	8	2	
11 ELECTRON LIBR	3	18	10	9	17	14	20	30	21	25	43
12 GOV INFORM Q			9	12	9	12	15	53	8	5	56
13 GOV PUBL REV	3		1			32	23	21	24	22	25
14 IFLA J-INT FED LIBR	4		1	14	33	11	25	22	29	18	
16 INFORM PROCESS MANAG	47	58	36	43	37	43	57	54	49	45	48
17 INFORM TECHNOL LIBR	28	89	38	32	36	22	37	18	20	26	38
18 INT FORUM INFORM DOC	14			17	21	22	23	18		6	

\7 SocialSci \81 LISA \144 Pascal \202 ISA

cross-database comparisons particularly informative, but cannot be appropri-
ately reproduced in black and white. For this reason the coverage charts show
only the pattern of the target database in this article.)

FINDINGS

Table 1 lists the total number of records found for each of the 42 journals in each
database. It also summarizes the total number of records by databases and provides
the subtotals by the three journal categories described earlier. Although *ISA* has the
longest time span (31 years) *Social SciSearch*, *LISA*, and *LibLit* outperform it in
the overall number of postings for the entire sample, and even for the *ISA* core
journal subset of category "A" titles. *Pascal*—with only 35 of the 42 journals
covered, and a 24-year time-span—had about 83% of ISA's postings. INSPEC had
46% of the postings of *ISA*, primarily because it covers only 34 of the 42 sample
titles, and of those only *Information Processing & Management* is considered as a
core journal by *INSPEC*. Also, as an engineering database, *INSPEC* is meant to
cover only information science- and technology-related articles from LIS journals.
   Table 2 identifies the major characteristics of the databases in the peer group,
and summarizes the number of records found for different variations of the 42 jour-
nal titles in each database. Results show that the position of *ISA* in the peer group
is further weakened when the search is limited to the past five years. Limiting the
search to the last year of the study period (1996) made the coverage in *INSPEC*

similar to that in *ISA*. This bird's eye view of the coverage reinforced the need for further exploration of the longitudinal pattern of coverage in *ISA*, the database believed to be the flagship database of the information industry. The other data-bases in the peer group could be as easily analyzed; however the focus here is on the coverage characteristics of *ISA* that shows so far the largest variety and gravest problems in the peer group. The analysis of the coverage of the individual journals across the entire time span of the database identified the following major problems in the journal coverage policy and practice of *ISA*: late start of coverage, early termination of coverage, temporary abandonment (gaps in coverage), tardiness of coverage, shallow coverage, uneven coverage and duplicate entries. As one journal may illustrate several deficiencies the coverage pattern graphs are grouped together. Figures 2 and 3 include the graphs for eight of the *ISA* core journals, and eight of the *ISA* non-core journals, respectively.

Late Start of Coverage

ISA has the longest retrospective coverage of the library and information science field, and for some journals in the sample set it fared well in terms of picking up journals since the first year of publication or the start year of the database time span. However, there are important exceptions. Of the core journals, *RQ* (that started in 1960 and is covered by *ISA* since 1970) and the *International Journal of Information Management* (that started in 1986 and is covered by *ISA* since 1989 with merely two records for that year) are late pickups. Worse is the picture when considering the early period of coverage not only the very first year. Though *ISA* covers the first two years of the *Annual Review of Information Science and Tech-nology*, its third and fourth volumes are not covered at all.

In the group of *ISA* non-core journals *Government Publications Review* starts on time but its second and third volumes are ignored, and the fourth one had a single record. (The journal split into two parts then merged again and carried the original title, but such difficulties and confusing changes were handled during the coding of journal titles retrieved by the loose searches in the JNF database). *ISA* picks up *IFLA Journal* on time but ignores it in the next five years, except for a single record in its third volume. *Library Acquisitions* appears with a single record from its first volume, but its next two volumes are ignored. One core journal, *Library & Information Science Research* (that changed its title in 1983 from *Library Research*), stood out for an impossibly early start with 8 records from 1980. The 8 records attributed to *Library & Information Science Research* actually appeared in the Japanese journal *Library and Information Science*, and were incorrectly iden-tified in the *ISA* records. The value of long retrospective coverage of 31 years somewhat diminishes when considering the citation half-life of LIS journals For most of the prominent LIS journals that the Institute for Scientific Information monitors, the yearly editions of *Journal Citation Reports* show a cited half-life between 2.5 and 6 years, with only a very few over 10 years cited half-life.

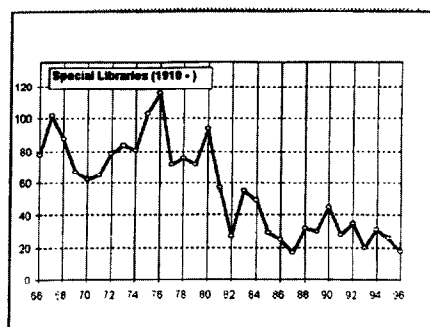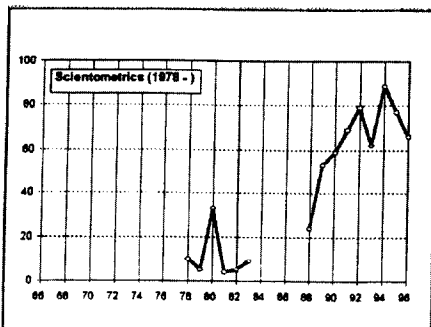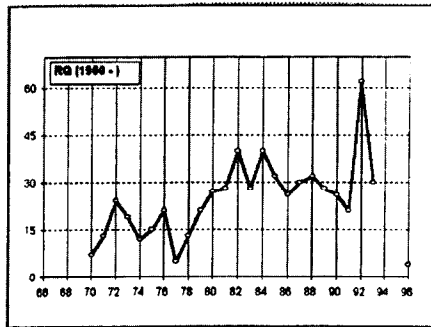**FIGURE 2**
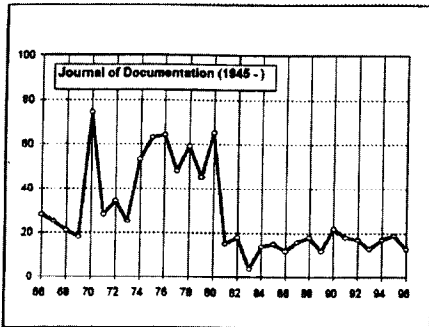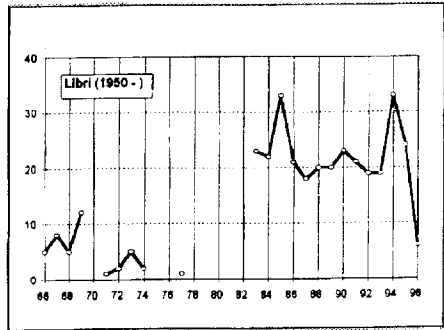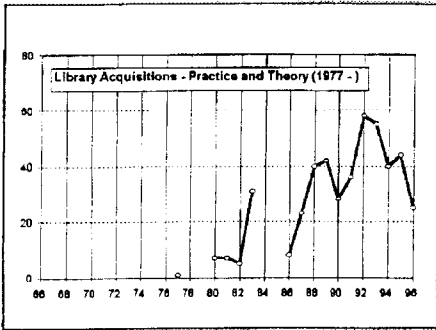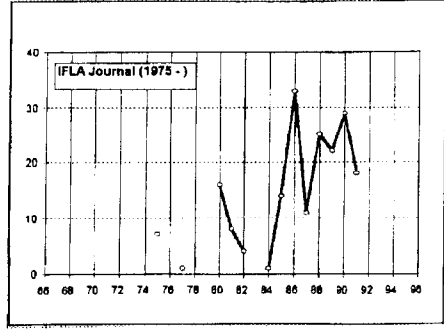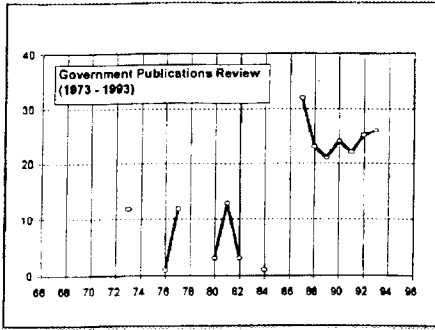**Coverage Pattern of Some ISA Core Journals**
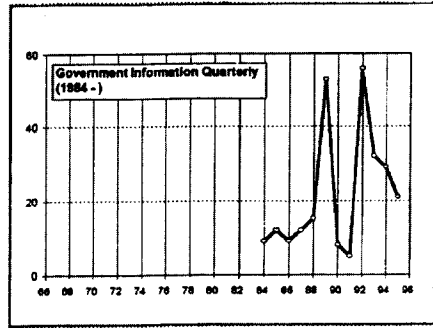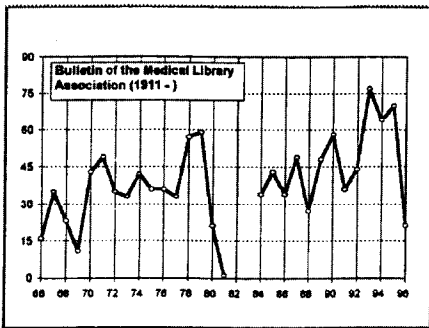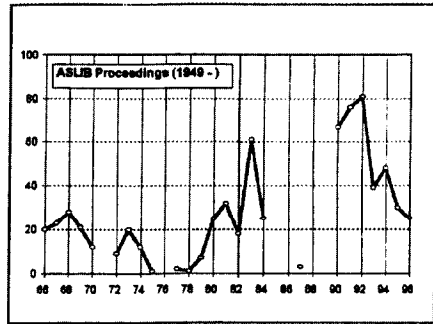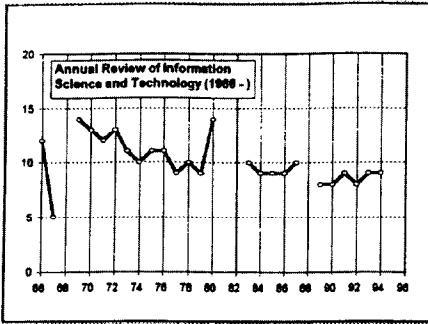
FIGURE 3
Coverage Pattern of Some Non-Core Journals
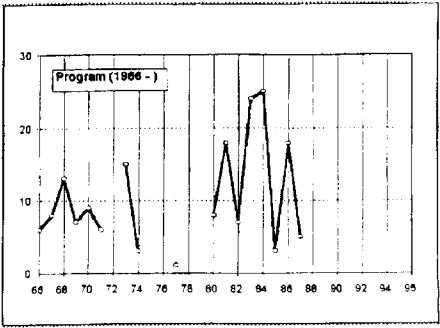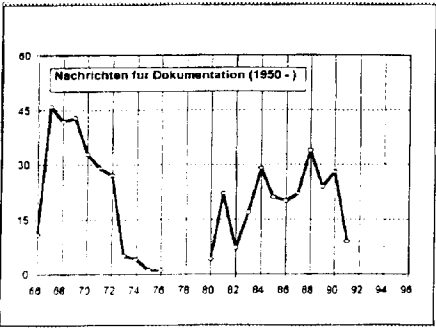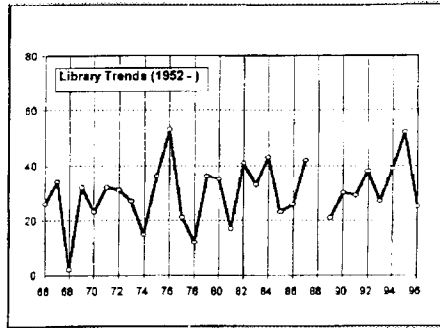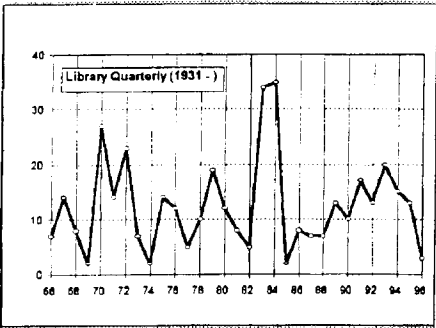
Early Termination of Coverage

Early termination is more of a concern than late start of coverage as the short citation half-life data clearly show users' preference for current publications. There is a high number of journals in the sample that were dropped from ISA despite their reputation in the field. Sometimes it is difficult to say if the coverage of a journal was terminated or it is "only" seriously delayed. Although for some core titles no records were added for the 1996 volumes (and for the *Annual Review of Information Science and Technology* not even for the 1995 volume) by January, 1997, the formal closing date of the study period, a follow-up search before the submission of the revised manuscript showed that these titles were not abandoned but only delayed, and belong to the category of tardy coverage.

As expected, most of the prematurely dropped titles were non-US, non-core journals in the sample (*Education for Information, IFLA Journal, International Forum on Information and Documentation, Nachricten für Dokumentation, Nauchno-Tekhnicheskaya Informatsya, Program, Scholarly Publishing*, and *Zeitschhrift für Bibliothekswesen*). This certainly weakened the international coverage claimed by *ISA*. The case of *Scholarly Publishing* covered between 1969 and 1984 is particularly strange as it was picked up again by ISA in 1994 when it changed its title to *Journal of Scholarly Publishing*. *Law Library Journal* that could illustrate every negative aspect of coverage in *ISA*, also had an abrupt ending just two years after *ISA* started to give it an acceptable level of coverage.

Only one core journal in our sample, *RQ* belongs to this category. *ISA* has no records after 1993 for articles in *RQ* except for four entries added in the last update of 1996. *RQ*'s impact factor hovers around 0.4 that makes it the 17th ranking among the 60 library and information science journals monitored by ISI in 1996. The ranking and overall reputation of the journal do not justify the termination of its coverage, Its resurfacing in the last 1996 update of *ISA* does not give enough hope for its phoenixing as only four records were added.

Temporary Abandonment (Gaps in Coverage)

Temporary abandonment differs from early termination because the journal reappears after one or more years of hiatus. It is probably worse than early termination of the coverage of a journal. The latter is an often arguable but at least a—presumably—deliberate decision. Temporary abandonment is mostly circumstantial and accidental. They are not acceptable for core titles, and can be hardly justified for other high prestige journals. In this regard both *ISA* core and non-core titles fare badly. The *Annual Review of Information Science and Technology, ASLIB Proceedings,* the *Bulletin of the Medical Library Association*, and *Scientometrics* have unacceptable gaps of several years, in some cases in a recurring fashion. Some of the titles with years of gaps in coverage are serials published by the sponsoring associations of *ISA*. It was hard to decide which category *RQ* would fit

better. Technically, it would belong to this category, because it did re-surface again in the last update of 1996. However, as only four records were added for the entire year of 1996, and none for 1994 and 1995 it is more likely to have been abandoned and was discussed under early termination.

Of the non-core *ISA* titles in the sample *Government Publications Review, IFLA Journal, Interlending and Document Supply, Library Acquisitions, Libri, Nachricten für Dokumentation, Program*, and *Serials Librarian* have gaps of several years.

Tardiness in Coverage

Sometimes the belated updates make it difficult to decide if a journal has been terminated, temporarily abandoned, or its records are just tardy. By the closing date of the study period no records were added for the 1996 issues of such core journals as *College & Research Libraries, Government Information Quarterly, Journal of Academic Librarianship, Journal of Education for Library and Information Science*, and the *Journal of Information Science*. There were no records added for the 1995 and 1996 volumes of the *Annual Review of Information Science and Technology*, a most essential tool for any information science research. Three *ISA* core journals, *Artificial Intelligence*, and two *IEEE Transactions* (not included in the sample discussed here) fared even worse, and seem to have been abandoned for good as of January, 1997. However, in a follow-up test search it was found that the 1997 updates added hundreds of records for articles published in these serials in 1993, 1994, 1995, and 1996. Such delays have characterized *ISA* since its beginning, and there have been no signs of substantial improvement. *ISA* had been criticized for lack of currency repeatedly. Bottle and Efthimiadis (1985) reported that "in all the previous studies *ISA* had the largest time delays," and "*ISA* was proved once more to be the poorest in timeliness."

Such delays are aggravated when the database itself is not updated according to the promised schedule as was the case repeatedly with *ISA* in the first 10 months in 1996 (Jacsó, 1997). The results of the final search at the closing of the study period clearly indicate that *ISA* has grave problems with the currency of its records. Table 1 shows that four of the five peer databases added more records for articles published in 1996 in the sample journals than did *ISA*. This was true even for the core journals of *ISA* except that in this category *ISA* outperformed Pascal marginally.

Shallow Coverage

Core journals are designated as such with the understanding that every item that is published in a journal will have a record in the database, unless otherwise noted. *The IFI Online Database User Guide* (1994) claims that "core journals are abstracted completely." There is no qualification, and no exclusion mentioned so

it can be assumed that records are created for all articles, brief communications, book reviews, and letters to the editors. The notion of complete coverage is reaffirmed in the 1996 *ISA* marketing pamphlet. It claims that "over fifty core titles are abstracted **in their entirety** [emphasis mine], including *Annual Review of Information Science and Technology, ..., Database, Electronic Library, Journal of Documentation.*" Even a cursory look at the postings information would make it clear that the coverage of *ISA* core journals is far from complete. This is, however, one of those measures of coverage that must be put in perspective by comparing it with data from one or more of the databases of the peer group. Data for each of the sample titles from *ISA* and the five primary peer databases in Table 1 show how often core journals have far fewer postings in *ISA* than in two or more of the other databases. Again, narrowing the time-span for this comparison to the last 5 years would further weaken the position of *ISA*.

College & Research Libraries, the journal with the highest impact factor among LIS journals is better covered by *LISA, LibLit,* and especially by *Social SciSearch* than by *ISA* (except for the fact that *ISA* goes back the farthest). This remains so even if book review records are excluded in the latter two databases that have exceptionally comprehensive coverage of book reviews.

In some cases when coverage by *ISA* seems to be competitive it often turns out to be attributable to duplicate and triplicate records to be discussed later. The majority of the other core publications, especially *Government Information Quarterly, Information Technology and Libraries, Journal of Academic Librarianship, Journal of Education for Library and Information Science,* and *Journal of Information Science* have particularly shallow coverage for their status. Other core journals such as *The Electronic Library, Database, Online,* and *RQ* are far from having cover to cover treatment.

Expectations for depth of coverage are lower for non-core titles, but such publications as *Government Publications Review, IFLA Journal, Interlending and Document Supply, Library Acquisitions, Library Journal, Library Trends,* and *Wilson Library Bulletin* would have deserved more comprehensive coverage at least in the years they were covered at all by *ISA. Program,* the excellent British journal for library automation that ranked 13th by its impact factor in the most recent *Journal Citation Reports,* was not only terminated very early, and completely ignored in some years, but it was also given shallow coverage during the few years while *ISA* included it.

Uneven Coverage

The coverage of a journal can be considered uneven when the number of records for a journal fluctuates by more than 20%-25% from one year to the other without a corresponding fluctuation in the number of items published in the journal. In the case of core journals this criteria is closely related to depth of coverage, though even a less than complete coverage is expected to be steady. As shown by the

roller-coaster shape of the graphs, *ISA* cannot meet even these more lenient criteria. Other databases also show considerable unevenness but not to the extremes of *ISA*. In some years the coverage drops to such a low level that it might as well qualify for a gap. This is the case, for example, in the *ISA* core journal category with *ASLIB Proceedings* in 1975, 1977, 1978, and 1987, *Journal of Documentation* in 1983, and *Scientometrics* in 1979, 1981, and 1982.

Journal of Academic Librarianship, Journal of ASIS, Online, and *Special Libraries* fluctuate more then the variance in the number of items published in these journals would justify. In the case of important but non-core journals selectivity is acceptable. However, heavy fluctuation in the number of entries across years is not. Still, such pattern was found with the majority of the sample titles, including the *Canadian Library Journal, Education for Information, Government Publication Review, IFLA Journal, International Forum of Information and Documentation, Library Acquisitions, Library Quarterly, Library Resources and Technical Services, Library Trends, Nachricten für Dokumentation, Program, Scholarly Publishing,* and *Serials Librarian.* As shown by the graphs in Figure 3, *ISA* non-core journals repeatedly hit bottom.

Duplicate Entries

For some titles the fluctuation is not as extreme as the year by year postings may show. On the other hand, neither is the coverage as extensive as the postings may suggest. These distortions are caused by the massive amount of duplicates in *ISA* that were found when checking the suspiciously high postings for some titles. After encountering duplicate and even triplicate records for the same articles in *ISA*, DIALOG's duplicate detection facility was used to determine the rate of duplicates for every journal in the sample in *ISA* and in the databases of the peer group. While DIALOG's algorithm is not perfect, its limitations did not handicap the databases, erring for non-duplicates in most of the cases. The false identification as duplicates of articles with non-distinctive titles ("Book Review," "Letter to the Editor," and "Annual Directory Issue") are more then offset by those real duplicates that are missed by DIALOG due to spelling variations and title/subtitle reversals in different records for the same article within a database. Even with this conservative approach, and discounting the false duplicates, the test searches in the sample found a 6% average duplication rate for the sample journals in *ISA*, and practically none in the other databases of the peer group. In *ISA* some of the sample titles showed above 10% rate of real duplication. Table 3 shows a partial list of duplicates and triplicates found for *Library & Information Science Research* that had 60 records in duplicates or triplicates out of the total of 294 records (subtracting 8 for the records mistakenly attributed by *ISA* to *Library & Information Science Research*). The evidence of and possible motivation for recycling earlier *ISA* records and the pattern of indiscriminate importing of redundant records from

TABLE 3
Sample Duplicates/Triplicates from *Library & Information Research*

Accession Number	Article Title (truncated if too long for this printout)
8701451	Availability studies in libraries.
8608951	Availability studies in libraries.
8701453	Barriers to library cooperation in Costa Rica.
8609044	Barriers to library cooperation in Costa Rica.
8603186	Comparing return rates of home loans of social science book material.
8602149	Comparing return rates of home loans of social science book material.
8701811	A conceptual framework for theory building in library and information science.
8608848	A conceptual framework for theory building in library and information science.
8603135	Development and evaluation of a measure of library automation.
8602094	Development and evaluation of a measure of library automation.
8608971	Effects of age, gender, college status, and computer experience on attitudes toward library computer systems.
8701452	Effects of age, gender, college status, and computer experience on attitudes toward library computer systems.
8508006	Eponyms and citations in the literature of psychology and mathematics.
8503316	Eponyms and citations in the literature of psychology and mathematics.
8600026	Evaluation of a career development and assessment center program for professional librarians.
8509059	Evaluation of a career development and assessment center program for professional librarians.
8510002	Implications for application of qualitative methods to library and information science research.
8509002	Implications for application of qualitative methods to library and information science research.
8502633	Innovation decision making and the genesis of OCLC. The effects of individuals, information, and structure.
8410700	Innovation decision making and the genesis of OCLC. The effects of individuals, information, and structure.
8408297	Innovation decision making and the genesis of OCLC. The effects of individuals, information, and structure.
8600116	Library books selected by elementary school students in Hawaii as indicated by school library circulated records.
8509141	Library books selected by elementary school students in Hawaii as indicated by school library circulated records.
8600254	Library and information science abstracting and indexing services: Coverage, overlap, and context.
8509058	Library and information science abstracting and indexing services: Coverage, overlap, and context.
8707357	Library and information science research. An analysis of the 1984 journal literature.
8800192	Library and information science research. An analysis of the 1984 journal literature.
8502475	The literacy education gap. The involvement of public libraries in literacy education.
8410271	The literacy education gap. The involvement of public libraries in literacy education.
8408496	The literacy education gap. The involvement of public libraries in literacy education.
8608063	Management and the conduct of in-house library research.
8603113	Management and the conduct of in-house library research.
8508226	Organization factors in multitype library networking.
8503105	Organization factors in multitype library networking.
8502129	Reference clientele and the reference transaction in five Illinois public libraries.
8410229	Reference clientele and the reference transaction in five Illinois public libraries.
8408190	Reference clientele and the reference transaction in five Illinois public libraries.
8410226	Reference effectiveness. A review of research.
8409140	Reference effectiveness. A review of research.
8408189	Reference effectiveness. A review of research.
8800193	A survey of children's librarians in Illinois public libraries.
8707358	A survey of children's librarians in Illinois public libraries.
8701469	Stratospheric aerosols. The transfer of scientific information.
8609301	Stratospheric aerosols. The transfer of scientific information.

other sources by IFI/Plenum are discussed by Jacsó (1997), and are the subject of further research. This unparalleled volume of duplicate and triplicate entries for the same items are disturbing in any kind of searches, and have detrimental effect on searches for bibliometric studies.

CONCLUSION

The analysis of the coverage of 42% of *ISA* core titles along with a group of high prestige LIS serials in six databases has shown deficiencies and questionable practices of the journal coverage. The most serious ones were found in *ISA*, the target database of the research. Although the selection of core journals in *ISA* is good (though one may wonder why the *Proceedings of the ASIS Annual Meeting*, and the *Bulletin of ASIS* are not considered core titles), in practice these core journals are not given the treatment they would deserve, and users are promised and expect. The coverage was found to be worse with category "B" titles that are ranked the highest by deans and faculty of LIS schools, and by directors of academic libraries. Significant and recurrent gaps, early termination of coverage, shallow and heavily fluctuating coverage were found in both the core and the non-core title groups. These deficiencies together with the unprecedented rate of duplicate records make the coverage of the database unpredictable, and inappropriate for comprehensive searches in the LIS literature, and especially for bibliometric studies.

The research methodology presented here can be applied to any database group in comparative evaluations of journal coverage. In this research the journal selection favored *ISA*, representing 42% of journals claimed to be core journals by *ISA*. A sample subset of journals matching the preferences of an individual researcher, a special user community, or the holdings of a library would provide a more level playing field. The methodology can be applied by those who regularly evaluate and compare databases, and make recommendations for the purchase and licensing of CD-ROM databases, or the selection of online databases for recurrently needed comprehensive searches.

REFERENCES

Bottle, Robert T., & Efthimiadis, Efthimis N. (1985). Library and information science literature: Authorship and growth patterns. *Journal of Information Science, 9*, 107–116.

Cronin, Blaise, & Pearson, Stephen. (1991). The export of ideas from information science. *Journal of Information Science, 16*, 381–391.

Duff, Alistair. (1995). The information society as paradigm: A bibliometric inquiry. *Journal of Information Science, 21*, 390–395.

Edwards, Tom. (1975). *A comparative analysis of the major abstracting and indexing services for library and information science.* Paris: Unesco.

Gilchrist, Alan. (1966). Documentation of documentation: A survey of leading abstracts services in documentation and an identification of key journals. *Aslib Proceedings, 18,* 62–80.

Goldstein, Sam. (1973). Statistical bibliography and library periodical literature—Part 4: 1972 abstracting, indexing, and contents coverage of library and information science periodicals. *CALL, 2,* 3–13.

IFI Online Database User Guide. (1994). Wilmington, DE: IFI/Plenum Data Corporation.

Jacsó, Péter. (1992). *Evaluation and selection of CD-ROM software, dataware and hardware.* Englewood, CO: Libraries Unlimited.

Jacsó, Péter. (1994). The hidden dimensions of databases. *Database, 17,* 40–44.

Jacsó, Péter. (1997). Péter's database picks and pans. *Database, 20,* 84–87.

Jacsó, Péter. (1998). The anatomy of a database. *Searcher, 6* (in press).

Kabir, S. (1994). Authorship trend and solo research in bibliometrics: a bibliometric study. *Library Science with a Slant to Documentation, 31,* 87–90

Kohl, David F., & Davis, Charles H. (1985). Ratings of journals by ARL library directors and deans of library and information science schools. *College & Research Libraries, 46,* 41–47

La Borie, Tim, & Halperin, Michael. (1981). The ERIC and LISA databases: How the sources of library science literature compare. *Database, 4,* 32–37.

La Borie, Tim, Halperin, Michael, & White, Howard D. (1985). Library and information science abstracting and indexing services: Coverage, overlap, and context. *Library & Information Science Research, 7,* 183–195.

Nicholls, Paul. (1989). Bibliometrics of the laserdisk application literature. *Laserdisk Professional, 2,* 106–109.

Pettygrew, Karen, & Nicholls, Paul. (1994). Publication patterns of LIS faculty from 1982-1992: Effects of doctoral programs. *Library & Information Science Research, 16,* 139–156.

Poyer, R. (1982). Chemical Abstracts coverage of the preclinical sciences journal literature. *Journal of Chemical Information and Computer Sciences, 22,* 5–8.

Schwartz, Candy. (1992). The CD-ROM journal literature: Where do you look? *CD-ROM Professional, 5,* 42–46.

Tjoumas, Renee, & Blake, Virgil L. P. (1992). Faculty perceptions of the professional journal literature: Quo Vadis? *Journal of Education for Library and Information Science, 33,* 173–194

White, Howard, D. (1990). Profiles of authors and journals in information science: Some trails of ORBIT's GET command. *Proceedings of the 11th National Online Meeting* (pp. 453–459). Medford, NJ: Information Today.

Wood, James I., Flanagan, Carol, &Kennedy, H. Edward. (1973). Overlap among the journal articles selected for coverage by BIOSIS, CAS, and EI. *Journal of the American Society for Information Science, 24,* 25–28.

Yerkey, Neil. (1993). Publishing in library and information science: Audience, subjects, affiliation, source, and format. *Library & Information Science Research, 15,* 165–184.