An Empirical Investigation of the Obsolescence Rate for Information Systems Literature

by Sally Jo Cunningham

An Empirical Investigation of the Obsolescence Rate
for Information Systems Literature

Sally Jo Cunningham
Dept. of Computer Science
University of Waikato
Private Bag 3105
Hamilton
New Zealand

email: sallyjo@waikato.ac.nz

Abstract: A synchronous study has been performed on four years of the International Conference on Information systems (ICIS) proceedings, to determine the obsolescence rate for the field of information systems (as reflected in the sub-topics covered by the ICIS conference). IS is found to have a relatively high obsolescence rate, similar to that of fields in engineering and the technology-dependent "hard" sciences. In addition, this study provides a categorization of the types of documents referenced by IS research, and presents an analysis of the obsolescence rate for these types. This type of categorization permits a finer-grained examination of patterns of information dissemination and use.

1. Introduction

Scholarly documents may be thought of as possessing a life-cycle: a document us "born" at publication, sees a productive life during the time span in which subsequently published articles reference it, and finally "dies" when it is not cited again. Many documents, of course, are still-born – studies indicate that between one-quarter [Koshy, 1976] and one-third of all scientific articles are never cited (Computer Horizons, 1976). Many others live only briefly: the average annual citation count for scientific documents that do manage to be cited is only 1.7 (Garfield, 1977a), and most scientific papers die within 10 years of publication [Price, 1965, as referenced in Diodato, 1993].

During this life-cycle, the usage that a document sees is not steady. Bibliometric studies have validated the common sense assumption that references to an individual document in a subject literature tend to decrease with the age of the document [see, for example, Diodato, 1993; Stinson, 1987]. Typically, documents receive their greatest number and frequency of citations immediately after publication, and the frequency of citation falls rapidly as time passes. The document is said to be "aging", and when it ceases to be cited it is "obsolete".

"Obsolescence" is a slightly misleading term, however, as it implies that the document's information has been discovered to be incorrect or no longer applicable. There are several other reasons why a document my cease to be cited. As a result becomes better known, it will appear in review articles that may be cited in place of the original publication. Eventually a document's contents may become so thoroughly incorporated into a field that its findings are held to be common knowledge and the document no longer cited: "obliteration by incorporation" [Garfield, 1977b]. Moreover, a researcher may not select the "best" work for citation, but instead often chooses documents on a basis of accessibility (including factors such as physical availability of a format, language, inclusion in indexing services, etc.) [Smith, 1981].

Finally, the characteristics of the users may be as important as the characteristics of a document in determining a document's rate of aging [Griffith, 1979]. For example, an active researcher in a field will likely require up to the minute materials; a textbook author will range back through earlier work to present more fundamental and generally accepted results to students; and a historian of science may reference texts that are centuries old and whose scientific findings have long been discredited.

The obsolescence rate of a *body* of literature, rather than that of a single document, is an estimate of the rate of aging for documents in that field. Different disciplines have different obsolescence rates, with the humanities and social sciences generally making longer active use of scholarly documents than do the sciences. Obsolescence studies are one tool for examining the patterns of information generation, dissemination, and use in a research community. Examination of past obsolescence rates allows us to chart general changes in the modes of scholarship over the years; evaluation of obsolescence studies rates for distinctive geographic regions and sub-topics will allow us to draw contrasts between research methods found in different countries and within different research cliques. A deeper understanding of the subject methods and how they change can only be beneficial in producing quality work in the future.

Studies on obsolescence have concentrated primarily on the physical and natural sciences, and to a lesser extent on the social sciences and humanities. The management, information systems, and information technology literatures have scarcely been examined by bibliometricians and scientometricians [representative studies include Evaristo, 1993; Prechelt, 1994; and Hall, 1985]. Few obsolescence studies have been conducted in these subjects [for example, see Cunningham and Bocock, 1995]. This study examines one aspect of the IS literature – specifically the International Conference on Information Systems (ICIS) conference proceedings – as a preliminary attempt to quantify the IS obsolescence rate. This study also categorizes the types of documents referenced by ICIS papers, and examines the obsolescence rates for these different formats.

2. Research Methodology

Two types of measures are in common use: synchronous obsolescence and diachronous obsolescence. For the latter metric, a sample of documents is selected from a point in the past, and the rate of citations to these documents are calculated on a year by year basis (generally through use of a citation index). Typically a "half life" is calculated as the time period in which half of all observed citations are received by the document set.

In contrast, a synchronous obsolescence rate is determined by proceeding from the present to the past: the publication dates of documents *referenced* by a sample set of documents are examined to determine how far back in time the more recent half of those references were published. This measure is called the median citation age.

Obviously the synchronous obsolescence rate is easier to calculate, since it requires access to the sample documents alone, without recourse to a citation index. Whether synchronous and diachronous obsolescence are statistically equivalent is debatable [Stinson, 1987; Gupta, 1990]; they are largely complementary, since secondary calculations for the two measures give different insights into the subject literature.

The information systems/information technology literature has a notoriously wide "spread" through computing, mathematical, management, science, and the social sciences, making it particularly difficult to trace citations through the relevant indices. For that reason, a synchronous methodology was chosen. The proceedings of the ICIS conference was chosen for analysis, as ICIS is a general IS conference that publishes a wide sampling of topics from the information systems field.

Technique for calculating median citation age

In this study the median citation age for an article is calculated using the discrete analysis method: publication years are treated as discrete units, rather than as a continuum of quarters, months, and weeks. To determine the median citation age for a single article, the publication dates of its references are first listed in reverse chronological order. For example, the reference dates for this paper are:

1995, 1994, 1993, 1993, 1990, 1989, 1988, 1988, 1987, 1986, 1985, 1982, 1981, 1979, 1979, 1977, 1977, 1976, 1976, 1970

The median year is 1985.5 (the average of the 10th and 11th of the 20 values). Subtracting 1985.5 from the year of publication yields a median citation age of 9.5 years (1995 - 1985.5).

To determine the median citation age for a given subject, this procedure is applied to a sampling of articles from the field. When the median citation age is calculated for more than one year of publication, the study is called *multisynchronous*.

Data collection

The selection of a body of literature and range of years as a focus for a bibliographic study is often dependent on pragmatic as well as theoretic factors—availability of documents, ease of access, etc. While these reasons played a part in the choice of subject for this study, the proceedings selected for analysis are also appropriate because of the prominence of the ICIS conference in the IS field and its wide coverage of IS topics.

This study analyzed a total of 3308 references from 102 articles, and calculated the median citation ages for the 1981, 1983, 1991, and 1993 ICIS proceedings (Table 1). Only full research articles were included in the study – short papers and panel discussions were excluded, as they have different referencing conventions. The multisynchronous methodology (covering four years, spread over a decade) provides a more statistically viable basis for evaluation, as well as pointing up changes in the referencing conventions over the period sutied.

	ICIS 1981	ICIS 1983	ICIS 1991	ICIS 1993
# of articles	22	18	32	30
total # of references	604	526	1045	1133
minimum # of references per article	4	13	6	6
maximum # of references per article	69	60	57	72
average # of references per article	27.45	29.22	32.65	37.77

Table 1. Number of articles and references in ICIS proceedings

Sources of bias in the method

At this point, it is appropriate to consider the forms of bias that may affect this study. Firstly, the references cited by a document may not necessarily be the an exhaustive list of the references actually used by the author. MacRoberts estimates that scientists reference no more than one-third of all relevant sources [MacRoberts, 1988]; an author may consciously choose to reduce an over-lengthy set of references by presenting a sampling of the relevant literature, may deliberately omit less scholarly sources of information, and may by convention omit citations to classic papers. These sources of bias cannot be detected by this study's methodology, and may skew the observed obsolescence rate. However, correcting for this type of bias is notoriously difficult, and involves re-constructing the research process through lengthy interviews with the source document authors. For the sake of practicality, most obsolescence studies choose the quantitative methodology used here.

Secondly, there is some controversy about how to select document sets for citation analysis. Selection criteria includes analyzing "core" serials of a topic [eg, Diodato, 1993], choosing a sampling of articles relating to a single topic but that are scattered over a number of journals [eg, Wallace, 1986], and performing a comprehensive study of a broad range of topics over a broad range of serials [eg, Line, 1979]. While the latter option is the most desirable for determining the obsolescence rate for a subject as a whole, the first two techniques are preferable in calculating obsolescence for a more narrowly focussed topic. For pragmatic reasons this study concentrates the ICIS proceedings, and acknowledges that selecting articles on this topic articles from a broader range of sources may give different results. The study is, however, multisynchronous, eliminating the bias that may occur when source documents are all published in a single year.

3. Results

The reference data from the ICIS proceedings was analyzed in three ways: to determine the types of document formats referenced by ICIS papers, to calculate an obsolescence rate for each of the four proceedings, and to determine an obsolescence rate for each document format that was referenced.

Analysis of reference document formats

The references for the ICIS articles were categorized as follows: single articles from serial publications (including journals, trade magazines, annuals, etc.); books or chapters in books; articles from conference proceedings; technical reports or working papers; unpublished dissertations; presentations at conferences or professional meetings; miscellaneous (including personal communications, raw data, publicity brochures, etc.); and references whose format could not be determined. Table 2 summarizes the type distribution for the ICIS proceedings studied. A chi-square test conducted on this data showed no difference among the distributions of reference types for the four years studied (chi-square = 74.55, df = 21, p < .0001). That these distributions remained stable over the decade studied is one indication that the referencing conventions of the field (as represented in the ICIS proceedings) are well-established.

reference type	1981 ICIS proceedings		1983 ICIS proceedings	
	number	percentage	number	percentage
serial	307	50.83%	259	49.24%
book	164	27.15%	158	30.04%
proceedings	43	7.12%	51	9.70%
technical report	36	5.96%	26	4.94%
dissertation	31	5.13%	11	2.09%
miscellaneous	9	1.49%	12	2.28%
presentation	12	1.99%	6	1.14%
unknown	2	0.33%	3	0.57%
totals	604	100.00%	526	100.00%

(a) 1981 and 1983 proceedings

reference type	1991 ICIS proceedings		1993 ICIS proceedings	
	number	percentage	number	percentage
serial	533	51.00%	652	57.60%
book	352	33.68%	350	30.92%
proceedings	74	7.08%	58	5.12%
technical report	40	3.83%	31	2.74%
dissertation	16	1.53%	19	1.68%
miscellaneous	18	1.72%	12	1.06%
presentation	11	1.05%	7	0.62%
unknown	1	0.10%	3	0.27%
totals	1045	100.00%	1132	100.00%

(b) 1991 and 1993 proceedings

Table 2: categorization of references

The dominant publication format is the serial for all four years studied. This result contrasts strongly with the situation in the arts and humanities, in which references to books greatly outnumber those to other types of source documents [Cullars, 1990]. That only approximately one-third of the ICIS references are to books appears to support the popular assumption that in the rapidly changing field of IS, books are likely to contain out of date material.

Obsolescence rate for the ICIS proceedings

A graph of the ages of articles referenced by the journals in this study is presented in Figure 1 (separate figures for each year are presented in the Appendix). The graph shows an exponential curve commonly found in obsolescence studies, and includes the initial climb described by Price as an "immediacy effect" [Price, 1970]. The references in the ICIS proceedings exhibit typical obsolescence patterns, without any obvious peculiarities.
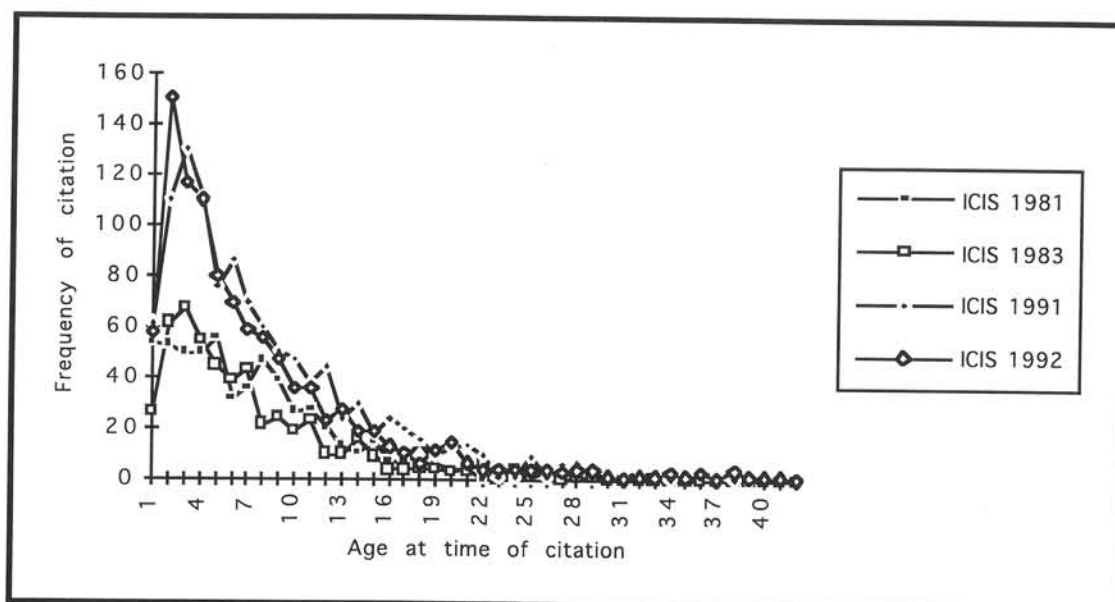
Figure 1. Distribution of ages of references found in ICIS proceedings

Table 3 presents a summary median, mean, and ranges of citation ages for the four ICIS proceedings. Note that the range of ages and the average citation age have increased slightly. This small "spread" in the distribution of reference ages has not affected the median citation age, however, which has remained stable at 5 years.

	ICIS 1981	ICIS 1983	ICIS 1991	ICIS 1993
median citation age	5	5	5	5
average citation age	6.97	6.52	7.38	7.73
# of articles	22	18	32	30
range of ages of references (years)	0 - 58	0 - 46	0 - 77	0 - 65

Table 3. Summary information of reference ages

What does a median citation age of 5 years imply about the obsolescence rate of the IS field? To place that number in perspective, consider the following table of median citation ages[1]:

---

[1]All studies are synchronous. With the exception the CS study, these results are taken from Diodato (1992). The CS result appears in Cunningham (1995).

Discipline	median citation age (years)
metallurgical engineering	3.9
genetics	4.0 - 6.0
operating systems (CS)	4.0
physics	4.4-4.9
chemical engineering	4.8
mechanical engineering	5.2
desalination	5.6
chemistry	8.1
archaeology	9.54
botany	10.0
music education	10.0 - 15.0
music theory	10.0 - 15.0
mathematics	10.5
geology	11.8
biblical criticism	21.63

Table 4. A sample of synchronous obsolescence results

As we might expect, the median citation age for IS is closest to values reported for engineering fields and other technology-dependent sciences. This result provides evidence for the oft-stated (but rarely empirically supported) belief that IS is a rapidly changing field.

Interestingly, the range in the publication dates is broader than might be expected, with citations to documents published as early as 1914 (for the 1993 proceedings). Inspection of the older references (pre-1950) reveal that they are primarily citations of classic results from other fields that are being applied to an IS problem (for example, psychological personality tests administered to users of computer systems). However, Figure 1 confirms that the publication dates of the references are indeed heavily skewed toward the publication date of the proceedings themselves (44% of the 1981, 49 of the 1983, 50% of the 1991, and 43% of the 1993 references date from the five years preceding the conference).

Obsolescence rates for different reference formats

The median citation ages for the various categories of reference document types are presented in Table 5. Note that references to proceedings and the "other" category (technical reports, dissertations, and oral presentations) are significantly "younger" than the serial references, and that in turn serial references are younger than book references. It appears that the "gray" literature tends to become obsolete most quickly – not surprisingly, given that these documents often contain preliminary results that are generally either published elsewhere or are discarded as they are discovered to be erroneous. That serials have a lower median citation age than books is also intuitively reasonable; journals tend to publish material closer to the frontiers of knowledge and to deal with narrower subjects, and therefore their results should be cited closer to their publication dates. Books, on the other hand, are more likely to present older results, perhaps as tutorials or overviews of a field.

document type	1981 ICIS med cit age	1983 ICIS med cit age	1991 ICIS med cit age	1993 ICIS med cit age
serial article	6	4	4	6
book	8	8	7	6.5
proceedings	2	2	2	2
all others	2.5	3	2	3

Table 5. Median citation age by document type

4. Conclusions

This study examines a commonly held belief about IS – that it is a rapidly changing field with a relatively high obsolescence rate for its documents. This hypothesis is confirmed, in that the median citation rate for IS is closer to that of engineering and the "hard" sciences than to the humanities and social sciences. Of course, IS is itself a broad term, and it is likely that the sub-fields of IS will have different obsolescence rates reflecting their historic origins or their "closeness" to other fields. This type of scatter has been measured in other disciplines [for example, physics; see Gupta, 1990]. Analysis of sub-fields of IS remains a topic for further research.

Comparison of the 1981/83 data with the 1991/1993 data shows that the obsolescence rates and types of documents referenced have remained steady over the decade. This result suggests that the referencing conventions for IS, as represented by the ICIS proceedings, are well established – an indication of stability for the field. Further, the obsolescence rate and distribution of citation rates is typical for a body of scientific literature. Again, this supports the existence of IS as an established field for which a consensus has developed about the appropriate types and ages of sources upon which researchers should draw.

The study goes beyond previous obsolescence research in its categorization of reference types and in its analysis of the median citation rates for these types. This categorization permits a finer-grained examination of patterns of information dissemination and use.
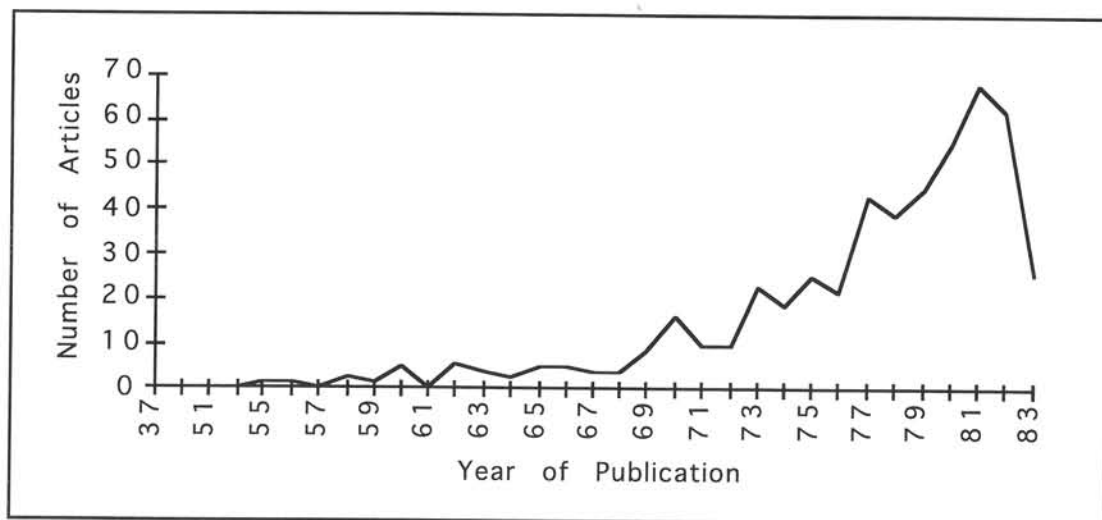
References

Computer Horizons, Inc. *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Reproduced by the US Dept. of Commerce National Technical Information Service, PB-252 339, 1976.

Cunningham, S.J., and Bocock, D. "Obsolescence of computing literature." To appear in *Scientometrics* (in press). Also available as *Working Paper Series 95/8*, Department of Computer Science, University of Waikato (Hamilton, New Zealand).

Cullars, J. "Citation characteristics of Italian and Spanish literary mongraphs." *Library Quarterly 60(4)*, 1990, pp. 337-356.

Diodato, V., and Smith, F. "Obsolescence of music literature." *Journal of the American Society for Information Science 44(2)*, 1993, pp. 101-112.

Evaristo, J.R., and Karahanna-Evaristo, E. "An empirical investigation of differences between North American and European MIS research." *Proceedings of the 14th International Conference on Information Systems* (Orlando, FL, USA), 1993, pp. 387-388.

Garfield, E. "Is the ratio between number of citations and publications cited a true constant?" *Essays of an Information Scientist,* vol. 2, ISI Press, 1977, pp. 419-421.

Garfield, E. "The 'obliteration phenomenon' in science – and the advantage of being obliterated!" *Essays of an Information Scientist,* vol. 2, ISI Press, 1977.

Griffith, B.C., Servi, P.N., Anker, A.L., and Drott, M.C. "The aging of scientific literature: A citation analysis." *Journal of Documentation 35(3),* 1979, pp. 179-196.

Gupta, U. "Obsoloscence of physics literature: exponential decrease of the density of citations to *Physical Review* articles with age." *Journal of the American Society for Information Science 41(4),* 1988, pp. 282-287.

Hall, R.P., and Kibler, D.F. "Differing methodological perspectives in artificial intelligence research." *AI Magazine 6,* 1985, pp. 166-167.

Khawam, J.J. "The AI interdisciplinary context: Single or multiple research bases?" *Library and Information Science Research 14,* 1982, pp. 57-74.

Koshy, G.P. "The citeability of a scientific paper." *Proceedings of the Northeast Regional Conference of the American Institute for Decision Sciences,* Philadelphia (PA, USA), 1976, pp. 224-227.

Line, M.B. "The influence of the type of sources used on the results of citation analyses." *Journal of Documentation 35(4),* 1979, pp. 265-284.

MacRoberts, M.H., and MacRoberts, B.R. "Author motivation for not citing influences: A methodological note." *Journal of the American Society for Information Science 39,* 1988, pp. 432-433.

McCain, K.W., and Turner, K. "Citation context analysis and aging patterns of journal articles in molecular genetics." *Scientometrics 17(1-2),* 1989, pp. 127-163.

Prechelt, L. "A study of experimental evaluations of neural network learning algorithms: current research practice." *Technical Report 19/94,* Fakultat fur Informatik, Universitat Karlsruhe, 76137 Karlsruhe, Germany, 1994. Anonymous ftp: /pub/papers/techreports/1994/1994-19.ps.Z on ftp.ira.uka.de.

Price, D. J. de Solla. "Citation measures of hard science, soft science, technology, and nonscience." Nelson, CE., and Pollock, D.K., eds. *Communication Among Scientists and Engineers.* Lexington, MA: Heath Lexington, 1970.

Smith, L. "Citation analysis." *Library Trends 30,* 1981, pp. 83-106.

Stinson, E.R., and Lancaster, F.W. "Synchronous and diachronous methods in the measurement of obsolescence by citation studies." *Journal of information Science 13,* 1987, pp. 65-74.

Wallace, D. "The relationship between journal productivity and obsolescence." *Journal of the American Society for Information Science 37(3),* 1986, pp. 136-145.
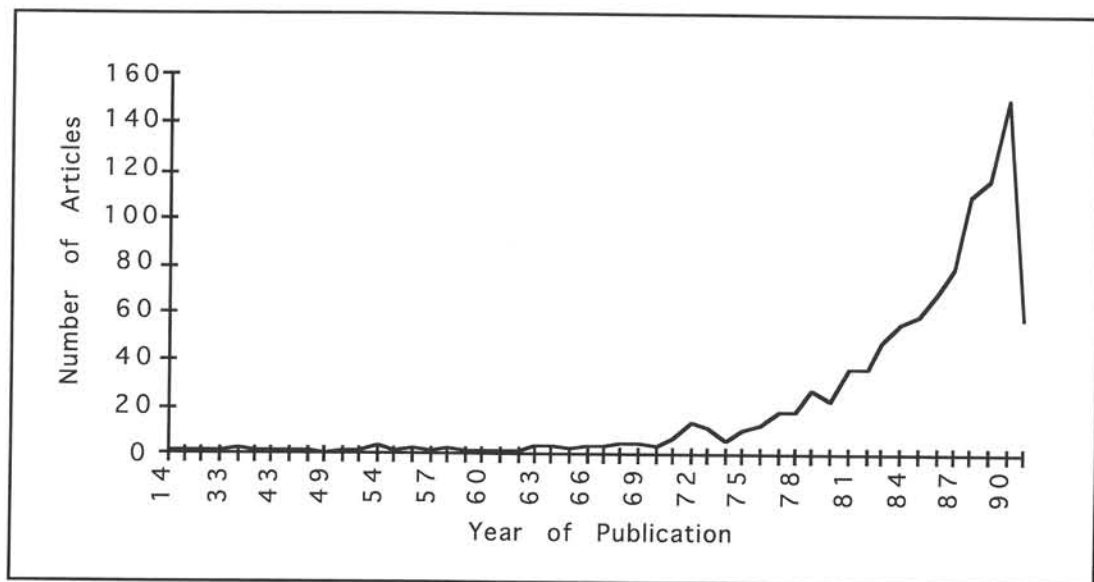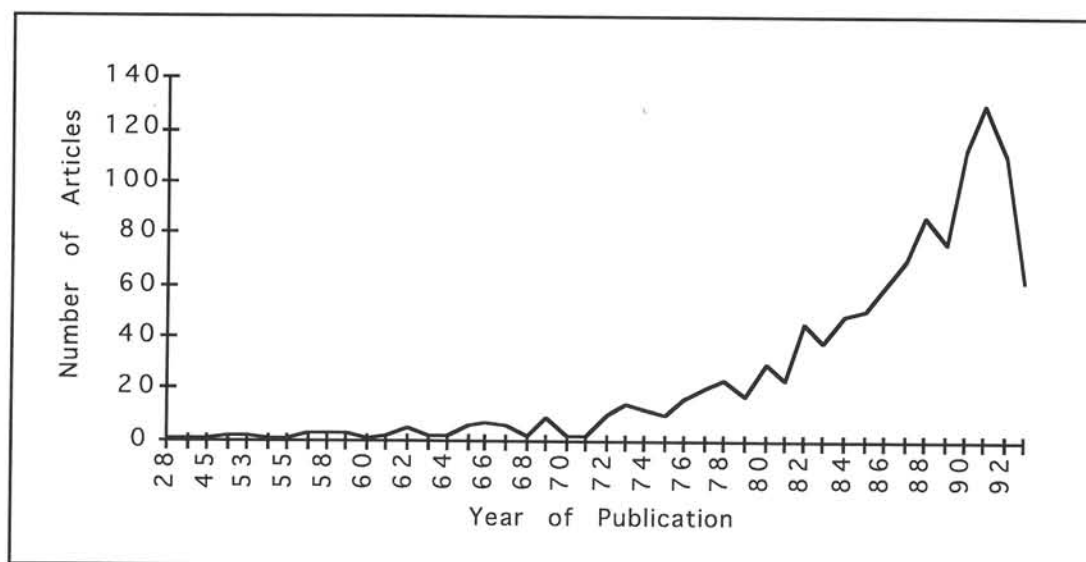
Appendix
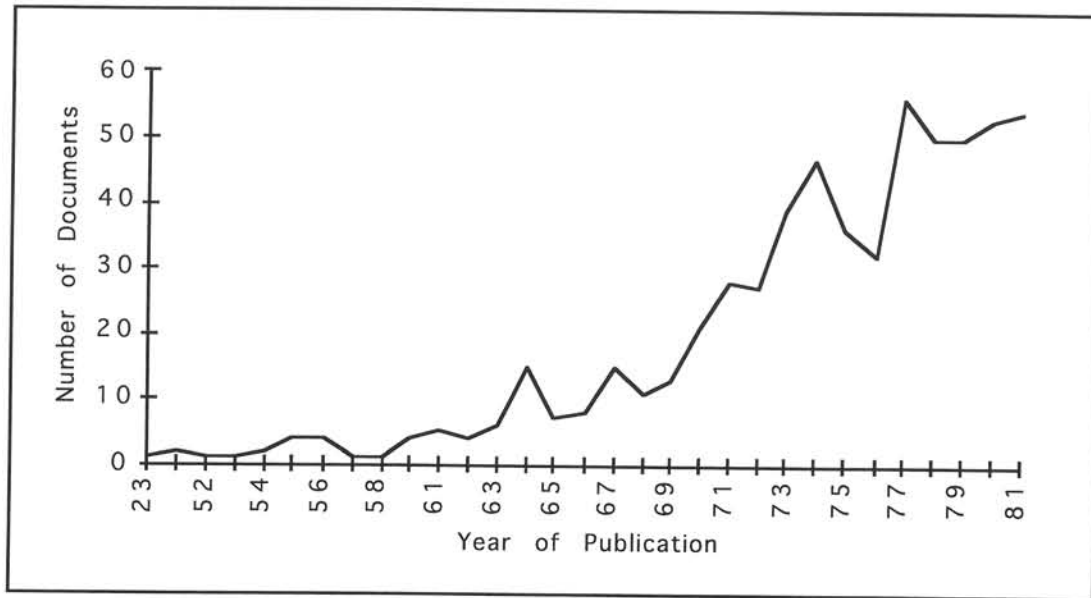
(a) ICIS 1981: number of articles referenced by year (1923-1981)

(b) ICIS 1983: number of articles referenced by year (1937-1983)

(c) ICIS 1991: number of articles referenced by year (1914 - 1991)

(d) ICIS 1993: number of articles referenced by year (1928 – 1993)