An Inquiry into the Domain of Information Science, with an Emphasis on Contributing Disciplines, 1973 to 1998: Preliminary Results

Boyd P. Holmes
Faculty of Information and Media Studies
University of Western Ontario
bholmes@julian.uwo.ca

Most information scientists appear to agree that the discipline absorbs, within its boundaries, all or part of certain other subject areas. Certain scholars, including Borko, Garrison, Rayward, James G. Williams and Martha E. Williams, have published, separately, what each of them considers those disciplines to be. My research, for which I present preliminary results in this paper, will, once completed, determine, through citation classification, which are the contributing disciplines of information science. The time period I will investigate will be 1973 to 1998, specifically the years 1973, 1978, 1983, 1988, 1993 and 1998. With each year, I am mapping a representative sample of citations, retrieved through Social SciSearch, against the Dewey Decimal and Library of Congress classification schemes. I include in this paper the preliminary results for 1973. Part of my research will also be the determination — again, bibliometrically — of those disciplines to which IS contributes, also for the period of 1973 to 1998.

A. INTRODUCTION

To define, Samuel Johnson informs us, is "to explain a thing by its qualities and circumstances." Arguably the most vexing problem facing information scientists today is their failure thus far to determine a consensable understanding of the boundaries of their discipline. Definitions of information science," wrote Alvin Schrader, correctly, in 1986, "are characterized by conceptual chaos." Three years earlier, he remarked, with equal astuteness, concerning the defining of both library science and information science, that a "rhetoric of labels rather than a logic of cognitive claims has prevailed." Since at least the start of this century, determining the limits of information science (IS) has been the focus of a body of academic research where, as with Schrader, scholars have executed exhaustive scutwork locating published explicit definitions for the discipline, then analyzed and catalogued their findings in the anticipation that defensible outlines for IS would then emerge. The most extensive of all previous investigations remains Schrader’s doctoral thesis (1983), a mammoth inquiry, in both breadth and depth, into explicitly-stated definitions of, not only information science, but also library science.

All previous research, however impressive much of it was, we may nonetheless consider unsuccessful because none of it produced a definition of information science that professionals and scholars subsequently accepted.
without question. Still, despite this failure, there appears to be a general agreement among most researchers that information science is an interdisciplinary subject area, one that absorbs, within its boundaries, all or part of certain other disciplines. (An exception among those scholars who consider information science interdisciplinary is Guy A. Marco, whose 1996 paper I mentioned earlier in this paper.)

In his classic essay “Information Science: What is It?” (1968), for example, Harold Borko enumerated those subject areas that, according to him, combine to create information science: communications, computer technology, the graphic arts, library science, linguistics, logic, management, mathematics, operations research, psychology, and “other similar fields.” Similarly, in her inaugural address as the president of the American Society for Information Science (ASIS) for 1988, Martha E. Williams provided this listing of the contributors to IS: classification science, cognitive science, communications, computer sciences, economics, electronics, information theory, library science, linguistics, logic, management science, mathematics, psychology and systems science. (More accurately, Williams considered those subject areas to be certain of the disciplines that combine to form IS.) For the moment, I am not concerned with the extent to which the enumerations by Borko and Williams were right (assuming that either or both of them were correct at all). I am instead just presenting two indications of how researchers concerned with the boundaries of information science consider it the summation of research from specific other disciplines. I chose these enumerations because of their serviceability in this regard; I selected Borko’s essay also because of its citation popularity among scholars.

In 1996, in the *Journal of Information Science*, Riitta Kärki produced this astute observation:

In order to understand research areas of scholarly disciplines, one should analyze not only the disciplines alone but also the relationships between different specialties. This concerns especially the most interdisciplinary fields. Information science has often been regarded as a good example of a specialty area that is remarkably interdisciplinary in nature, but there is not very much empirical evidence either to support or to refute this assumption.

I will offer that evidence. The subject of my research, for which I give preliminary results below, is the determination of which disciplines have contributed to the knowledge base of information science during the past quarter-century. Note that empirical determination of the knowledge base for IS is logistically feasible starting only with the year 1973. I explain below the reason for this limitation. My research will also be concerned with determining those subject areas to which information science contributes, also for the previous quarter-century.

Why does my inquiry matter? In 1992, Christine L. Borgman and Ronald E. Rice published, in the *Journal of the American Society for Information Science (JASIS)*, an inquiry into the extent of scholarly convergence, from 1972 to 1991, between communication and information science. Borgman and Rice’s justification for their inquiry is equally applicable to any investigation of the disciplines that contribute to IS and those disciplines to which IS contributes. The authors hypothesized: (1) if research...
determines that convergences between disciplines (in their case, communication and IS) are occurring, professors in those subject areas involved should consequently use a broader range of scholarship in both their teaching and research; (2) alumni of those subject areas that researchers discover to be converging could use the information to expand their employment possibilities; (3) faculty appointments committees might use the new knowledge in the hiring of scholars; and (4) the selection decisions of both collection development librarians and of editors of indexing and abstracting journals might improve as a result of the determination of new disciplinary boundaries. 13 It is self-evident that Borgman and Rice are right.

B. OVERVIEW OF PREVIOUS SCHOLARSHIP

i. Qualitative Inquiries

Despite the failure of information science thus far to determine its desired consensile understanding of itself, there is, as Schrader discovered, a plethora of published definitions for the domain, and theorists have – since 1981, when Schrader finished the data collection for his thesis14 – continued to publish articulations of the domain of IS. These persons included Emilia Currás (1985), Robert M. Hayes (1985), Richard A.V. Diener (1989), L.M. Bohnert (1989), Michael Gorman (1990), P.J. Lor (1992), Guy Marco (1996), W. Boyd Rayward (1996) and James G. Williams (1998).15

"The interdisciplinarity of information science," wrote Rayward in his 1996 article, "is a continuing theme in attempts to define it."16 That is true. There are, however, at least two problems with all existing writings on the scope of information science, faults my own research will eschew. These flaws stand apart from the fact that IS has chosen not to accept any of the definitions the prior scholarship advanced.

The first problem with most of the inquiries is that they have concerned themselves solely with the analysis of explicit definitions for information science. This method is flawed, since it fails to consider those definitions of the discipline that are implicit. One of the hypotheses of my research is that what information science claims to be differs from what it understands itself to be. In 1998, Howard D. White and Katherine W. McCain, presenting the results of their inquiry using author co-citation analysis of journal literature to map IS scholarship for the years 1972 to 1995, observed:

In our view, most of the authors mapped here are united in working on aspects of literatures as modes of communication. The intellectual problems they address arise from a particular kind of information system – literatures as content-bearing objects – rather than from “information systems” in general. Literatures are the distinguishing mark of domains likely to interest domain analysts, and at least a part of their interest in literatures lies in features that are exploitable for retrieval. The partitioning of literatures on the basis of such features, so as to most usefully approximate [sic] an answer to a request, defies the retrievalists’ enterprise.17

White and McCain then analyzed the descriptions of information science by Hayes and Borko I cited above. The analysis is worth quoting:

86
The rich word "information" has seduced some into characterizations of their field that are, to date, overgeneral. Thus, ACA [author co-citation analysis] can help assess their pronouncements in the disciplinary press. It leads to a judgment of "highly unfocused" on definitions of the field such as those quoted in Rayward (1996, p. 4): "The application of information science results in an information system. The role of information science is to explicate the conceptual and methodological foundations on which existing systems are based" (Borko, 1968, p. 67). Or "Information science is the study of the means by which organised structures (which we call 'information systems') process recorded symbols to meet their defined objectives" (Hayes, 1985, p. 174). These definitions would have ASIS-style information science dealing with employers' payrolls, a housewife's grocery receipts, Nightline, arrival and departure listings in airports, the Grand Ole Opry, and color-coded vial caps used by crack dealers. Need we say that it does not?

White and McCain, that the implicit comprehension of the scope of information science will prove to be much narrower than most of the explicit definitions thus far advanced. An investigation into clock design during the Italian Renaissance, for example, does fit the definition of information science that Hayes proposed. But information scientists would not consider such research as falling within the borders of IS, as proven by their failure to cite such research in their own scholarship (and the lack of such subject matter in the information-science curricula that they implement in their schools). On the basis of preliminary readings of explicit definitions of information science, I suspect that the sole type of case where the implicit definition might not prove smaller than, and contained by, the explicit definition, will be when the stated definition includes a chronological restriction: for example, Rayward, in his 1996 article, defined information science as a "term that is now conventionally used for attempts within the last 50 years or so to study in a formal and rigorous way processes, techniques, conditions, and effects that are entailed in improving the efficacy of information, variously defined and understood, as deployed and used for a range of purposes related to individual, social and organizational needs [emphasis added]." Later in that paper, Rayward did call for a broader conception of the chronological scope of IS, but the argument he gave is here beside the point: judging from his definition, he would claim that IS would not now accept, as being within the parameters of its concerns, seventeenth-century clock engineering.

The second problem with prior writings on the definition of information science (IS) concerns the published listings of disciplines that contribute to IS. I have already cited two such listings, by Borko and Martha E. Williams; to those
add listings by Guy Garrison, Rayward, and James G. Williams. For both convenience and future reference, I have gathered all five enumerations into one table (Table 1). It is on the next page.
<table>
<thead>
<tr>
<th>SCHOLAR</th>
<th>PROPOSED DISCIPLINES</th>
</tr>
</thead>
</table>
The five listings in Table 1, like all previous published enumerations, are flawed, in that their respective authors presented them without any justifications for why they were in fact the subject areas chosen: there is no proof of their presence in information science. My research, once completed, will determine conclusively which disciplines contribute to IS. I will also discover by their absence which subject areas do not contribute to IS, regardless of any unsubstantiated, and hence invalid, claims by specific scholars to the contrary.

ii. Citation Analyses

My actual inquiry into the disciplines that form the interdisciplinary will be a citation analysis. Borgman and Rice, in their article, cited above, provided sound justifications for using bibliometrics in the study of scholarly communication. Those justifications are as follows: (1) conclusions based on bibliometrics are reliable since other scholars may replicate the results; (2) the usually great size of most bibliometric datasets means the method allows easily for the mapping of longitudinal trends in subject areas; and (3) citation analysis is practical in achieving an overview of past and existing processes in communication among scholars. All these arguments confirm the validity of my selecting the bibliometric method as the means to answer my research question. In this regard my methodology will be unique. It is true that several other researchers into the definitional perimeters of IS, including Borgman and Rice, did use citation analysis as their primary methodology. However, these researchers executed their analyses in ways that I consider inadequate as a means to define the discipline in any fully acceptable way. I provide brief outlines of these studies directly below.

In 1981, Henry Small published, in Information Processing and Management, the first quantitative inquiry into the shape of information science to use author co-citation analysis, specifically a single-link cluster analysis using data from the Social Sciences Citation Index for the years 1975 to 1977.23

Also in 1981, Howard D. White and Belver C. Griffith published, in the Journal of the American Society for Information Science, “Author Cocitation: A Literature Measure of Intellectual Structure,” another author co-citation analysis study in information science, in this case for the years 1972 to 1979.24

In 1981, too, Bluma C. Peritz published, in Library Research, her inquiry “Citation Characteristics in Library Science: Some Further Results from a Bibliometric Survey.”25 Peritz’s matched 4,596 citations, from the years 1950, 1960, 1965, 1970 and 1975, to one of twelve categories she herself had predetermined. Peritz’s methodology, for the part of her research with which I am concerned, was problematic. Her classification scheme was questionable: since it is not a universal classification system, such as Dewey, Peritz’s system fails to allow for the emergence of subject areas the researcher might not have expected to appear. Universal schemes also have the virtue of permitting far greater detail of classification, and, as is the case with Dewey, the presentation of differing levels of detail. (As I will demonstrate, this strength of Dewey is one of which I will take full advantage in my research design.) Finally, Peritz’s subject was library science, not information science. Though this difference is not a problem as far as the quality of her research is concerned, it
does compromise the usefulness of her scholarship regarding my own inquiry.


In 1992, the *Journal of the American Society for Information Science* issued "The Convergence of Information Science and Communication: A Bibliometric Analysis." This is the article by Borgman and Rice that I cited twice previously; as I wrote earlier, the authors tried, in their inquiry, to determine whether the disciplines of communication and IS were, as the commonality of many of their research subjects suggested, indeed converging. Borgman and Rice's inquiry involved the examination of cross-citation patterns between the two subject areas.

Two years after Borgman and Rice, in 1994, Olle Persson published, also in the *Journal of the American Society for Information Science*, "The Intellectual Base and Research Fronts of JAIS[J] 1986-1990." For his inquiry Persson used, as had Small, single-link clustering to create clusters of the co-cited authors, to which he then applied multidimensional scaling to create a map of the intellectual base for IS. Persson then confronted, as the second major component of his research, the question of mapping shared areas of scholarly research, including the various respective proximities of those areas to each other.

I interpret Persson's research with caution because of a significant limitation to his methodology: its narrow focus. We cannot map a discipline with any accuracy using just one of its organs of publication, regardless of the significance of that organ within that community. Persson's selected chronological unit of analysis also restricts the potential usefulness of his study: because that unit included just five consecutive years, it cannot be said to measure longitudinal alterations in disciplinary foci.

In 1996, Riitta Kärki published her essay, from which I quoted above, "Searching for Bridges between Disciplines: An Author Co-Citation Analysis on the Research into Scholarly Communication" in the *Journal of Information Science*. For this article, Kärki used author co-citation analysis to map both IS and the sociology of science, for the period of 1972 to 1991, to determine the extent to which research into scholarly communication acts as a link between the two subject areas.

Finally, and most recently, in this cataloging of attempts to define the boundaries of IS, White and McCain published their 1998 article, from which I have already quoted, "Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995," in the *Journal of the American Society for Information Science*. White and McCain's title encapsulates the focus of their paper.

All of the above papers, regardless of their thoughtful conclusions, are limited as mappings of IS as a discipline, largely because of specific methodological flaws. The most significant of these errors was, on the part of Small, White and Griffith, Saito, Persson, Kärki, and White and McCain, the use of author co-citation analysis. Author co-citation analysis is inadequate as a means to determine domain because, by its nature, it concerns itself with determining just the core of a specific discipline. Thus, author co-citation analysis provides the reader with only a partial understanding of the scope of a subject. For my research, I am interested, not only in disciplines and sub-disciplines at the core of a subject area,
but also those areas of inquiry that the domain considers as being just somewhat within its concerns. Responding adequately to this interest will (1) allow for a more complete determination of domain, and (2) display potential shifts in the borders of the discipline: alterations that, because of their respective characteristics, might well occur close to the periphery of the discipline, rather than at its core.

White and McCain, and Kärki, in their respective papers, raised explicitly another problem with author co-citation analysis: it will have required considerable time for the authors being mapped to establish themselves sufficiently for their citation counts to be great enough for consideration. Kärki added that, once an author is indeed so situated, it takes more time for the cited research passes into the data collection of any co-citation study. Author co-citation analysis thus fails to map recent developments in the foci of a subject area.

Certain of the papers I addressed also included their own, specific limitations, which I gave in my respective descriptions of those papers. White and McCain's citation analysis contained an additional fault: the authors presented their source list of the twelve citing serials selected for their inquiry was without any detailed justifications for their being chosen.

The age of prior quantitative research into the domain of information science is also problematic. Time has made many of the prior bibliometric inquiries less useful than they were when published: Small's study stopped with the year 1977, White and Griffith's first study with 1979, Saito's with 1983, Borgman and Rice's with 1987, Persson's with 1990, and Kärki's with 1991.

The extent to which previous inquiries are dated I illustrate in the following figure (which spacing constraints force me to present over the course of two pages). The years from 1950 to 1998 run consecutively in grids across the top and bottom of the figure. This figure is divided into two parts: the first part consumes the top two-thirds of the pages, the second the bottom third. In the case of the top part of the figure, the years mapped in the eight horizontal bars are those the quantitative studies investigated; the names of the author or authors of the inquiry in question I give either immediately to the left or the right of the bar. I use one or two shades with each bar: a square with a heavy shade represents a year that the inquiry in question investigated directly, while a year with a light shade represents a year that the study investigated by implication. The bottom part of the figure maps, by year, the total number of inquiries into each year, both those studies that investigated IS, for the year, directly (the heavy shade used in the upper part of the figure) and indirectly (the light shade).
Figure 1: Years Investigated in Previous Quantitative Analyses into Domain of Information Science

Peritz (1981), continues
Figure 1: Years Investigated in Previous Quantitative Analyses into Domain of Information Science

| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 3 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

Small (1981)

White and Griffith (1981)

Peritz (1981), continued

Saito (1984)

Borgman and Rice (1992)

Persson (1994)

Kärki (1996)

White and McCain (1998)
Two interesting facts emerge from the bottom-section of the figure: (1) a sharp upswing in studies in 1972, a reflection of the fact that indexing of Social SciSearch started its indexing with that year, and (2) a tapering off in quantitative inquiries shortly after the start of the 1990s. The fact that part of my study investigates 1993 and 1998 directly, and 1990 to 1992 and 1994 to 1997 by implication, makes my inquiry one of the few to map highly recent developments in the shape of information science.

C. METHODOLOGY

My research divides itself into three parts: which I call Part A, Part B and Part C, respectively. Of these three sections, I have thus far completed just part of Part B. Part A will be qualitative, and Part B is quantitative; outlines of those two parts follow. Part C will provide a synopsis of the results of the thesis research, and map pathways for future scholarly inquiries. I cannot, of course, having not completed Parts A and B, predict the contents of Part C.

i. PART A

With Part A, I am concerned with assembling all enumerations, both partial and complete, of the disciplines that combine to form IS. These will include those listings by Borko, Garrison, Rayward, James G. Williams and Martha E. Williams (see Table 1). Chronologically, my research for Part A will start with 1968, the publication year of Borko’s essay. The selection of any earlier start date would be inappropriate, since the distance between that year and 1973, the first year for which I will be analyzing data, would be so great that any comparison of results for 1973 to pre-1968 enumerations would lack sufficient meaning.

ii. PART B

For Part B, my concern will be information science as it defines itself implicitly. I am using a bibliometric analysis to determine this definition, though my tool is different from the one Small, White and Griffith, Saito, Persson, Karki, and White and McCain all chose, which was, as noted, author co-citation analysis.

I divide Part B into two components.

First Component

For the first component of Part B, I am determining the disciplines of those publications that information-science literature has cited. To this end, I am executing a bibliometric analysis on a list I created of IS serials. To create the list I first located the titles of all the serials that the Social Science Citation Index (SSCI) classifies, in its 1997 Lists of Source Publications, as “Information Science & Library Science.” To eliminate those library science serials that were not also IS serials, I then mapped the SSCI list against the Library of Congress Subject Headings (LCSH), retaining those journals that the LCSH recognized as IS publications; specifically, I determined which records contained LCSH that started with either “Information science,” or “Documentation” – the latter term being, of course, an alternative term for IS, one no longer, in North America, commonly used – and ended with “Periodicals.”

This entire process gave me a list of sixteen IS serials indexed by Social SciSearch, four of which, I soon discovered, the database also indexes under respective earlier titles, and one of
which \emph{(Nachrichten für Dokumentation)} it now also indexes under a new title \emph{(NFD Information: Wissenschaft und Praxis)}. (Note that the \emph{List} does not indicate previous versions of citing serials.)

On the next two pages is a representation (Figure 2a) that displays, by year, the sixteen serials under consideration, with each bar representing one of the journals. There are three possible shades for each bar. The lightest of the three indicates that the serial in question did exist during the period in question, but Social SciSearch did not then index it; a heavier shade indicates that the serials was indexed, in whole or in part, for the year in question, but this is a year I am not considering in my inquiry; the heaviest shade indicates a serial that was both indexed on Social SciSearch for the year in question, and is under consideration in my inquiry. Note that variant versions of a serial are listed together, with the older version directly on top of the newer one.

On the third page following is a second representation (Figure 2b) that displays the totals, per year, of all the three types of shading, as they collapse against the same chronological grid as the first two pages. Totals for each of the three shading-types shadings run at the top of the columns; note that the number of citing serials that I am investigating in my research (darkest shade) are, by each of the six years under consideration, four (for 1973), eight (1978), eleven (1983), twelve (1988), thirteen (1993) and fifteen (1998). (I explain below why I selected these particular years.) Totals, by year, for the number of serials mapped in the two figures, run immediately beneath the chronological grid at the bottom.

In Figure 2b we see an overall growth, from the 1970s to the 1980s and 1990s, in the number of IS serial publications; this trend reflects the expansion in our discipline. Because of this greater quantity, I will have more data, and hence more dependable results, for the 1980s and 1990, than for the 1970s.
Figure 2a: Citing Serials Used in Inquiry: 1973 to 1998

- Canadian Journal of Information Science
- Canadian Journal of Information and Library Science
- The Information Society
- Information Processing and Management
- The Information Scientist
- Journal of Information Science
- Information Technology and Libraries
- International Forum on Information and Documentation
- The International Information and Library Review
- Journal of Academic Librarianship
My research for Part B is proceeding as follows. I am limited in my inquiry to scholarship published after 1971, since, as I wrote earlier, Social SciSearch indexes social science serials only for the years 1972 to the present. There exists no other database, online or in print, with the requisite data for any year earlier than 1972.\textsuperscript{35} As I illustrated in Figures 2a and 2b, the years I select for study are 1973, 1978, 1983, 1988, 1993 and 1998. I chose those years carefully. They are equidistant from each other, thus preventing, as would happen were each year not spaced in such a manner from the other, the slanting of my analysis in the direction of any specific periods. The five-year gap between the years allows for differences between the periods to be seen with clarity; changes in citation patterns are thus made stark, whereas a blurring between areas would had I instead picked two years, or ranges, either close or next to each other. Also, the arrangement of years allows the greatest possible reach into the past (back to 1973) while remaining as close as possible to the present (1998, the full year closest to when I will execute my data collection, 1999). With each search on Social SciSearch, I will isolate the data by year, and by the sixteen serials being considered. I then further limited my search, this time by document type (dt=article). (Note that Social SciSearch does not use the this delimiter for any records for 1972.) I selected only articles since none of the remaining document types indexed are useful to me in my inquiry: this is because, as I wrote on earlier, I am interested in how researchers implicitly understand information science to be; and published research in the discipline is almost invariably published in prose works where the delimiter is “article.”\textsuperscript{36} What is more, as Persson wrote concerning his methodology, the restriction of retrieved SciSearch data (in his case, citations to JASIS) to only articles ensures that most of the subsequently retrieved items do indeed contain references.\textsuperscript{37}

Once my set is formed, I use, in each case, the rank command (rank cw) to retrieve my data for cited works in ranked order. The downloaded citations I am now mapping using two universal classification schemes; details of this process I give below. My analysis of the resulting data is focusing on the determination of those disciplines information-science scholars cited for the period under investigation.

Two possible protests to my methodology may be raised at this point. One objection is that I ignore references in IS research that was published in monographs. The inclusion of citations from books will not, however, affect my data substantially. Empirical research by Kathleen Garland (1991), and Karen E. Pettigrew and Paul T. Nicholls (1994), on scholarly publishing within library and information science (LIS), gave proof that LIS scholars usually publish their work in serials, and not in, or as, monographs.\textsuperscript{38} Because of that research I anticipate that my citation analysis of IS literature will confirm that most scholarly communication, within the interdisciplinary, is through journals, rather than book. (Thus far it has, as I will indicate later below.)

Of course, my methodology also restricts me, regarding my citing IS documents, to articles published in those IS serials that Social SciSearch indexes as Information Science and Library Science. However, my criterion for selection has provided me with a body of citing serials that, according to the references retrieved in a pilot study I ran for 1990, form the majority of those publications that IS cites the most. The additional possibility that those IS serials not indexed by Social SciSearch maintain their own,
idiosyncratic patterns of citation, is one I may need to investigate, as part of my research for Part B.

That pilot study for 1990 made several additional discoveries, all of which indicated that the process of executing this component of Part B would be complicated. To start with, use of the rank cw command on Dialog File 7 ranks an item by the number of times an article cites it at least once, but not necessarily just once. What is more, it is possible for different items to be collapsed into a single abbreviation: a reference to B.C. Vickery’s *Classification and Indexing in Science* (second edition, 1959), and one to D.J. Foskett’s *Classification and Indexing in the Social Sciences* (first edition, 1963), for example, Institute for Scientific Information (ISI) technology files together under the abbreviation CLASSIFICATION INDEX. Because of these two problems, I need to download, using the view command, each of the ranks I am investigating to determine both (1) the actual number of times an item has been cited within that rank, and (2) the items that are included in that rank.

There are several additional difficulties with this section of the thesis research. To begin with, the size of the data sets, for each year, necessitates sampling: once I reach the level of two-citations-per-rank, I will examine, for each ten rankings, just the tenth citation rank, extrapolating the resulting information to each of the respective remaining nine ranks.

Another problem is that ISI provides the user with only *abbreviations* to the items being cited – the company does not also include the titles themselves. ISI also fails to maintain a key to its abbreviations for this function: such notations ISI generates only automatically. A further complication regarding this part of my research is that a download of any citations does not in fact reveal the identity of the acronym in question. Rather, it shows only the context in which the acronym appears. Thus, my prototype study revealed Virgil Diodato’s article “The Use of English-Language in Non-United-States Journals – A Case Study of Mathematics Publications, 1970-1985,” published in *Library and Information Science Research*, in 1990, cited the following work: “WOOD DN, 1967, V23, P117, J DOC”. A glance at the numbering sequence for the volumes of *The Journal of Documentation*, relative to the specific years in which the respective volumes were published, shows that this citation is indeed almost certainly to the serial in question. I therefore interpreted the reference, here, as being to *The Journal of Documentation*. With the decipherment of the more obscure references, I have been consulting the citing documents in question.

The process through which I am classifying the citations I describe now. The most convenient way to locate appropriate classifications for the cited articles is to use an easily accessible classification scheme. In this regard, I am locating, for all cited items under study, their respective Dewey Decimal and Library of Congress (LC) classification notations. In the case of serials, those are usually determined using the *Ulrich’s International Periodicals Directory*. With monographs, one of the best sources for such notations is the online catalogue of the University of Toronto, known as UTCat; this catalogue is my source for classification notations for books. UTCat is accessible through the Internet, and, though the library itself uses only LC for its own classifications, Dewey classifications are usually available through the MARC records of the catalogue, records that are accessible to the online user. What is more, UTCat includes records for over eight million volumes, making the likelihood great that I will be
able to retrieve through the catalogue
most of the monograph notations I will
need.\textsuperscript{40} \textsuperscript{41}

As part of my analysis, I am
\textit{graphing} the results for my Dewey and LC
mappings. Dewey, with its divisions into
tens, hundreds, and thousands, lends itself
with particular ease to the creation of
illustrative visual representations,
specifically column graphs.

Part of my analysis for this section
of Part B also consists of determining, for
the six years under investigation, the exact
percentage of the citations specifically to
information-science literature. This will
show me, by comparison, the \textit{extent to}
which IS has absorbed information from
other disciplines, regardless of what those
subject areas may be.

The entire process through which
the data for the first component of Part B
will be compiled and analysed has, at this
early stage of the research, already
involved a substantial expenditure in time
and intellectual effort. I cannot, however,
justify a simpler model. I am concerned
that any less complex approach – one, for
example, where far fewer journals are
examined for their citation counts –
would compromise to an unacceptable
point the integrity of whatever
conclusions to my research I will reach. It
is in the nature of citation analysis to
approximate: we cannot infer that an
academic with a lifetime citation count of
one hundred was more accomplished a
scholar that a colleague with a count of
just eighty citations, nor can we
consequently claim that the second
scholar had exactly eighty percent of the
importance of his colleague. We can,
however, conclude that both academics
almost certainly had more impact on
scholarship than the professor with a
lifetime count of just ten references.
When used appropriately and responsibly,
citation analysis gives us an admittedly
imprecise, but nonetheless valid, image of
the use and transfer of scholarly
knowledge. Given that I will, at times, be
sampling from the data collected, and
thus limiting the detail of my resulting
portrait of the discipline, a simpler
bibliometric analysis than the one I have
selected would consequently make the
yielded quantitative image of IS indistinct
to the point of incomprehensibility –
Jackson Pollock in place of Edward
Hopper.

\textbf{Second Component}

For the second component of
Part B, I will execute another series of
citation analyses. In this case, however,
those analyses will not be of information-
science literature, but rather of the
respective literatures of specific other
disciplines, my intent being the
delineation of the patterns for those
subject areas regarding the extent to
which each of them has, from 1973 to
1998, used information-science literature.
This is the \textit{self-derivation factor}, which is the
extent to which a discipline contributes to
other disciplines.\textsuperscript{42} The years chosen will
be the same six as the ones for the first
component, for the same reasons I gave
on earlier. It is with this component, in
particular, that my research for Part A will
inform my inquires for Part B. The
disciplines I will investigate be those
subject areas that specific scholars have
identified as contributors to or invaders
of IS; this disciplinary list will of course
include those enumerations listed in Table
1. Among the subject areas I will
investigate, for this sub-component, will
be library science (LS), since part of my
research interest includes the
determination of the relationship between
LS and IS. I will also investigate the
citation patterns of those additional
subject areas that the first component for
Part B will have proven to have

102
contributed, for the period under consideration, to information science.

This component of Part B will involve a somewhat different methodology than did the first component of the same Part. The methodology for this second component is as follows. For each of the disciplines that my research for Part A and the earlier component of Part B will have suggested might be encroaching, or have encroached, on information science, I will determine, through the most recent published annual ISI Subject Category Listings of the Social Sciences Citation Index, Arts and Humanities Citation Index and Science Citation Index, the most commonly cited journals for that area. The exact number of serials I will consider for each subject area I cannot give without first having completed the research necessary to produce the list of disciplines to be investigated; with those disciplines for which ISI indexes few serials, I will include all serials in the analysis, while with those disciplines, such as computer science, for which ISI indexes many serials, I will analyze the top ten journals, for that discipline, as the citation impact factors for the most recent issues of the Listings indicate. (I choose the number ten arbitrarily: I submit that ten serials will be sufficiently large a sampling of data to be generalized to each population in question.) In the case of library science, I will use the Library of Congress Subject Headings to determine my list of LS serials. I will not, however, include in that list any of the serials that the Library also classifies, through its Subject Headings, as information-science journals, since those will be among the serials mapped in Figure 2a, the citations of which I will already have investigated as part of my data analysis for the first component of Part B.

Upon the determination of the serials under consideration, I will, for each discipline, execute a series of citation analyses similar to the ones for the first part of Part B. However, with this component, there will be, relative to the previous component, four differences in my methodology. First, I will be searching for only references to information-science publications, since citations to works belonging to other disciplines will not yield data relevant to my research question. (Most of the possible ISI acronyms for the information-science serials I will have by then determined through my analysis, for the earlier component of Part B, on the citations in IS.) Second, I will not be matching references against classification notations, since I am not with this research concerned with the determination of the complete mapping of these disciplines. Third, I will not be downloading the entire bibliographies of the respective articles to which the individual citations belong; although the examination of the frequency of the citation of information-science publications within each article would yield more precise data than the approach I plan to take, the work involved in achieving such accuracy would fall outside the prescribed time frame my research imposes. Fourth, and finally, there is, consequent to the third exception just noted, no need, with this component of the quantitative aspect of my research, for any sampling after I reach the ranks of two citing articles: with this section of the thesis research, I will examine all downloaded citations.
D. PRELIMINARY RESULTS

Thus far all my results are for 1973 only. For that year I investigated four serials (Information Storage and Retrieval, the Journal of the American Society for Information Science, the Journal of Documentation and Nachrichten für Dokumentation; cf. Figure 2a). Until the two-citations per rank level was reached, there were seventy-one ranks; once I investigated each rank, and separated those references that ISI software had, because of similarities in titles, collapsed into the same acronym, the number of rankings stood at 101. Note that I have not yet investigated those citation after the three-citations-per-rank level.

Further investigation into all the rankings revealed a total, until the start of the two-citations-per-rank level, of 652 citations, of which 543 I have thus far classified, seventeen of which are unclassifiable, and ninety-two remain to be classified.

Of the 543 classified citations, 493, or approximately ninety-one percent, are to serials, and fifty, or approximately nine percent, are to monographs. This predominance of references to serials confirms, at least thus far in my research, the conclusions of (1) the two studies by Garland, and Pettigrew and Nicholls, and (2) my pilot study, that journal publication is the principal vehicle of the communication of research among library-and-information-science scholars. In making this finding, I am also confirming the defensibility of my decision to derive my citations entirely from serials, rather than from monographs too.

I repeat the above findings in Tables 2a, 2b and 2c, on the next page.
Tables 2a-2c: 1973 Totals, Until Two-Citations-Per-Rank Level

Table 2a: Number of Citations Classified

<table>
<thead>
<tr>
<th>Classified Citations</th>
<th>543</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassifiable Citations</td>
<td>17</td>
</tr>
<tr>
<td>To Be Classified</td>
<td>92</td>
</tr>
<tr>
<td>Total</td>
<td>652</td>
</tr>
</tbody>
</table>

Table 2b: Unclassifiable Citations Until Two-Citations-Per-Rank Level

<table>
<thead>
<tr>
<th>Citations to Theses</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations to Personal Communications</td>
<td>4</td>
</tr>
<tr>
<td>Internal Document of Limited Circulation</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 2c: Classified Citations by Document Type

<table>
<thead>
<tr>
<th>Classified Citations</th>
<th>Number of Classified Citations</th>
<th>Percentage of Classified Citations, to Nearest Hundredth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serials</td>
<td>493</td>
<td>90.79</td>
</tr>
<tr>
<td>Monographs</td>
<td>50</td>
<td>9.21</td>
</tr>
<tr>
<td>Totals</td>
<td>543</td>
<td>100</td>
</tr>
</tbody>
</table>
I have thus far determined the LC and Dewey classification notations for the main classes of the schemes only. I present on the next two pages the results for these mappings, as Figure 3 (LC) and Figure 4 (Dewey), respectively. Note that these figures were created just before certain data were mapped; the totals at the bottom bars of the two figures (559 for LC and 553 for Dewey) therefore do not match exactly the data for Table 2a (543 classified + 17 unclassifiable = 560).

In the case of Figure 3, the viewer's first observation is the extent to which IS cited itself: just over 77% of the citations thus far mapped were to Z (library science, which, in the LC scheme, includes information science). There was little spread, among the classes, of the remaining citations: just five of the remaining twenty main classes mapped at all in this regard, and of those, with four of the classes the number of citations was too small to be of any significance. Only the main class for science (Q) enjoyed a substantial number of citations, at almost 15 percent of the references mapped.

With Figure 4, Dewey indicated the same level of self-citation for IS in 1973 as did LC: by far the greatest number of citations (83%) were to the 0 class, which includes library and information science (specifically the 020s). Of the nine remaining classes, five mapped. Of those five, just two mapped substantially: Natural Sciences and Mathematics (class 5), at approximately 8.5 percent, and Technology (Applied Sciences) (class 6), at approximately 3 percent.

A thorough mapping of the citations for 1973, which I hope to complete in time for the presentation of this research at CAIS in June 1999, will, in particular, allow me to judge empirically the accuracy Borko's 1968 enumeration of the contributing disciplines to information science (Table 1). Recall that,
Figure 3: 1973: Information Science Citations, Until Two-Citations-Per-Rank Level, Classified According to the Library of Congress Main Classes

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>Count</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Works</td>
<td>A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Philosophy, Psychology, Religion</td>
<td>B</td>
<td>0.89 (5)</td>
<td></td>
</tr>
<tr>
<td>Auxiliary Sciences of History</td>
<td>C</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>History: General and Old World</td>
<td>D</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>History: America</td>
<td>E</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>History: America</td>
<td>F</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Geography, Anthropology, Recreation</td>
<td>G</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Social Sciences</td>
<td>H</td>
<td>2.33 (13)</td>
<td></td>
</tr>
<tr>
<td>Political Science</td>
<td>J</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Law</td>
<td>K</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>L</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Music and Books on Music</td>
<td>M</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Fine Arts</td>
<td>N</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Language and Literature</td>
<td>P</td>
<td>0.54 (3)</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>Q</td>
<td>14.67 (82)</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td>R</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>S</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>T</td>
<td>1.43 (8)</td>
<td></td>
</tr>
<tr>
<td>Military Science</td>
<td>U</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Naval Science</td>
<td>V</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Library Science</td>
<td>Z</td>
<td>77.1 (431)</td>
<td></td>
</tr>
<tr>
<td>Unclassifiable</td>
<td></td>
<td>3.04 (17)</td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td>100 (559)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4: 1973: Information Science Citations, Until Two-Citations-Per-Rank Level, Classified According to the DDC Main Classes

<table>
<thead>
<tr>
<th>Category</th>
<th>Citations</th>
<th>Value</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalities</td>
<td>83</td>
<td>0.9</td>
<td>459</td>
</tr>
<tr>
<td>Philosophy and Psychology</td>
<td>1</td>
<td>0.9</td>
<td>5</td>
</tr>
<tr>
<td>Religion</td>
<td></td>
<td>0.54</td>
<td>3</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>3</td>
<td>0.72</td>
<td>4</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td>0.54</td>
<td>3</td>
</tr>
<tr>
<td>Natural Sciences and Mathematics</td>
<td>5</td>
<td>8.5</td>
<td>47</td>
</tr>
<tr>
<td>Technology (Applied Sciences)</td>
<td></td>
<td>3.25</td>
<td>18</td>
</tr>
<tr>
<td>The Arts; Fine and Decorative Arts</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Literature and Rhetoric</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Geography and History</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unclassifiable</td>
<td></td>
<td>3.07</td>
<td>17</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>99.98</strong></td>
<td></td>
<td><strong>553</strong></td>
</tr>
</tbody>
</table>
NOTES

1Samuel Johnson, A Dictionary of the English Language: in Which the Words are Deduced from Their Origin and Illustrated in Their Different Significations by Examples from the Best Writers. To Which are Prefixed A History of the Language, and An English Grammar, 1 (London: "Printed for G. and J.," 1822), 485.

2I take the adjective "consensible" from Alvin Schrader (Alvin M[arvin] Schrader, Toward a Theory of Library and Information Science [PhD thesis] [Bloomington, Indiana]: Indiana University, 1983, 1) who in turn took it from J.M. Ziman's book Public Knowledge: An Essay Concerning the Social Dimensions of Space (Cambridge, England: Cambridge University Press, 1968), 11. The word is a neologism used to identify "rational opinion about which there exists collective scholarly agreement [op. cit., Schrader, "Toward," 1]." For "opinion," I prefer the word "judgement."


5In his chapter "Definitions about Information Science [ibid., 94-204]," Schrader wrote, "Almost 700 definitions of information science and its antecedents are examined in this chapter from the English-language press between 1900 and 1981 [ibid., 99]." It is interesting to note that the oldest reference in Schrader's bibliography for this chapter is a citation to a French-language essay Paul Otlet published in 1903 (ibid., 956), while what appears to be the oldest reference to an English-language publication is to one from 1939, specifically S.C. Bradford's "The Coordination of Documentation" (ibid., 934). The bibliography also contains reference to a document, or to documents, from 1937 (ibid., 970), specifically Preliminary Reports for the World Congress of Universal Documentation of that year; I do not as of yet know the language or languages in which these Reports were published.


10Ibid., 18.
As proof of that appeal, consider this: I executed an online citation search on the article, using the online versions of the Institute for Scientific Information (ISI) indexes Arts and Humanities Search (file 439 on Dialog), SciSearch (file 434), and Social SciSearch (file 439); note that Social SciSearch starts indexing on Dialog only with 1972, Arts and Humanities Search with 1980, and SciSearch with 1974. The search produced a citation count that, once duplicate I removed duplicate entries, stood at twenty-five citations, six of which were from articles published in the 1990s. The essay is also often cited in monographs, most recently in Richard E. Rubin’s textbook *Foundations of Library and Information Science* ([New York: Neal-Schuman, 1998], 407).

I executed the search on 5 January 1999.


Ibid.


Ibid. 14-5.


Henry Small, “The Relationship of Information Science to the Social
Howard D. White and Belver C. Griffith, “Author Co-citation: A Literature Measure of Intellectual Structure,” *Journal of the American Society for Information Science*, May 1981, 32(1):63-71. I credit Small with being the first to use author co-citation analysis in this regard because his paper was received on 3 April 1980, and White and Griffith’s paper on 14 May 1980; see *ibid.*, [163]; *op. cit.*, Small, 39.


Peritz also recorded, in this article, discoveries concerning both the citations she identified, and the research methodologies of the papers to which those citations referred. Though her findings, and her analyses concerning them, are all intriguing, none of them are useful to my research, and I therefore do not describe them here.


All the definitions I locate will be in English. (French is the only other language of which I have an adequate understanding, specifically an acceptable reading knowledge.) The language restriction of this research will probably not bias it extensively, given that virtually all inquiries into the matter have been published in English.


The creation of such a tool is also beyond the time available to me for the completion of this thesis. Note, too, that White and Griffith in 1981, Kärki in 1996, and White and McCain in 1998, all, because of the same limitation in existing data, started their respective inquiries with the year 1972 (cf. Figure 1). Though I am critical of her work in other respects, I nonetheless salute Peritz for having found the time to execute citation analyses on library-science literature published in the years 1950, 1960, and 1965.

According to the expand command on Dialog, the remaining delimiters are: Abstract; Art Exhibit Review; Bibliography; Book Review; Chronology; Correction; Addition; Dance Performance Review; Database Review; Discussion; Editorial; Editorial Material; Excerpt; Fiction, Creative Prose; Film Review; Hardware Review; Item about an Individual; Letter; Meeting Abstract; Music Performance Review; Music Score Review; News Item; Note; Poetry; Record
Review; Reprint; Review; Review; Bibliography; Software Review; Review; Radio Review; Video.


40 University of Toronto School of Graduate Studies 1998-1999 Calendar (Toronto: School of Graduate Studies, University of Toronto, 1998), 98.

41 The Library of Congress online catalogue might seem a more appropriate choice for the source of my classification notations, since its decisions regarding the classifications of specific items are influential within the community of North American libraries. However, I find the Library of Congress online catalogues, compared to UTCat, impractical as tools for the collection of my data: three out of the four LC catalogues are inaccessible during certain specified hours during the week; MARC records do not exist for all entries, making the determination of Dewey notations impossible for certain items; and, in my experience, the online network for the site often fails to initialize. The fact no catalogue, among the four available, indexes all LC items, further inhibits the usefulness of the LC online catalogues to my research. In contrast, UTCat indexes all its documents in one catalogue, is always accessible, does provide MARC records for its items, and only rarely becomes inoperative.


BIBLIOGRAPHY


University of Toronto School of Graduate Studies 1998-1999 Calendar. Toronto: School of Graduate Studies, University of Toronto, 1998.


