

CHAPTER 3. THE FORMATION OF A LITERATURE COLLECTION ON BRADFORD'S LAW OF SCATTERING

Introduction. In this chapter, a collection of documents is formed on the research topic of Bradford's Law of Scattering, using the general procedure developed in the previous Chapter. The collection formed is precisely defined, compatible with conventional interpretation of the topic, and comprehensive.

Outline of Procedure. The procedure proposed may be outlined as follows:

- One begins with a notion of the topic that is sufficient to identify: (a) A&I Services with the smallest subject scopes necessary to cover the 'centre' of the topic but with wide temporal and geographic coverage of the scholarly literature; and (b) in these A&I Services, probable good index terms for the topic. From the A&I Services, one obtains a set of documents which form a correct seed for the formation of the collection. A careful analysis of these documents should provide some understanding of core notions of the topic, and suggest an initial working selection criterion for its literature. Further, from bibliographic properties of this sample, it should be possible to infer the broad outlines of the topic in terms of a time-frame, the geography of its publication, and so on; this will guide the planning of a search strategy.
- To expand the collection from the seed, an extensive search of the literature is undertaken, using a variety of search tactics in parallel, for example, the use of other A&I Services and the tracing of bibliographic references. The search is terminated when the return, that is new documents judged to be on the topic, is negligible vis-à-vis the number of unsuitable documents scanned.
- Of the new documents retrieved in searching, those judged to have any likelihood of being on the topic are inspected. Using the working selection criterion, documents are selected as 'most likely on' the topic, or are rejected. The former add to previously acquired documents with this classification. The latter are reconsidered in terms of the prior understanding of the topic, and those judged to be possibly incorrectly rejected are placed into a 'possibly on' or 'marginal' category with any such previously assigned documents. Most documents in the remaining largest category of 'most unlikely to be on' the topic are discarded, though a number may be retained to provide background samples.
- Periodically, all documents acquired to that time in the two retained collections are carefully studied, and the topic and its boundaries reappraised. This may produce a

slightly different understanding of the topic and a slightly different working selection criterion for documents subsequently retrieved. It will also require a re-classification of the existing collections. The original seed literature, as a part of the existing collections, could also be reassessed, though any but the smallest of changes here would indicate a deviation in the perception of the topic from the conventional view. Drastic revisions would also require a repetition of earlier searches to recover the discarded literature. In due course, well before the termination of the search, re-assessments should produce no further modifications either to an understanding of the topic or to the selection criterion.

- At this stage, the selection criterion is tested for its rejection of documents in the background. Failure to adequately reject background documents would require a revision of the selection criterion, and, possibly, a reappraisal of the earlier understanding of the topic and adjustments to the collections.

File of Documents. Appropriate data on each suitable document were recorded in electronic files. Documents most likely on the topic were placed in what is termed the 'BRAD File', while all other documents, for example those marginal to the topic, or those citing or cited by each BRAD File document, were placed in what is termed the 'Non-BRAD File'. The BRAD File records represent not only analytical-level scholarly documents, which are the principal focus of the present work, but all documents discovered which meet the selection criterion, even those such as dissertations which were not systematically searched for. Each *record* in a file represents one distinct document and consists of a set of *fields*, labelled with a numeric code, for example (1) DOCNUM, which contains a unique record number for each document, or (16) YEAR, which contains the publication year for the document. For documents on the topic as well as for documents marginal to the topic, a hard copy of each document is separately stored, allowing for later analysis and the creation of new fields. Files were stored in three different computer systems over the period of the project, viz. CYBER 72, IBM 3090 and 486 PC; they were manipulated with a variety of system-provided software and with numerous customised programs in versions of BASIC appropriate to the system. Files, and their composition at the closure of the collection, and manipulation tools are detailed in the Appendix.

Outline of Chapter. In the description of the implementation of the procedure which follows, details of each intermediate collection assessment are not given; only the final assessment on the completed collection is fully described. The procedure is presented in six Sections:

1. The Recognition of the Topic by A&I Services and the Formation of a Literature Seed.

2. A Search for Documents on the Topic.
3. An Analysis of the Subject Content of the Literature on the Topic. (One component of the proposed graphical selection criterion for the topic literature is also determined).
4. An Analysis of the Marginal Literature. (A detailed examination is made, and the final status determined, of documents lying on the boundaries of the collection).
5. An Analysis of the Form of the Text of the Topic Literature. (This allows the proposed graphical selection criterion to be completed with a second component).
6. The Rejection of the Background by the Selection Criterion. (The immunity of the proposed selection criterion to deception by homographic usages of its component words in the background literature is checked; a satisfactory result completes the criterion).

A seventh section generalises key aspects of the above procedure.

1. RECOGNITION OF TOPIC BY A&I SERVICES, AND FORMATION OF LITERATURE SEED

A first step to forming a collection of documents on the topic of Bradford's Law of Scattering is to obtain a sample of documents which are recognised as on the topic by suitable A&I Services. Three A&I Services which specialise in Library and Information Science -- in which the topic is centred -- were chosen for this purpose: *Library Literature* (1921-), *Library and Information Science Abstracts* or *LISA* (1969-), and *Information Science Abstracts* or *ISA* (1966-). Only indexer-created fields, representing the classificatory judgements of A&I Service experts, were searched. Author-provided fields, such as the Title and (usually) the Abstract, were ignored. All index terms containing the word-stem '-BRADFORD-' were inspected, and those compatible with the present author's 'primitive notion' of the topic were selected. Terms deemed unsuitable from the outset are described in appropriate footnotes, and terms deemed suitable are fully detailed below.

Library Literature in printed form was searched manually for terms with this word-stem in the alphabetic Author/Subject index of each volume from 1921. *LISA* and *ISA* in electronic form were searched online through the DIALOG system for occurrences of the word-stem in the Descriptor field. As the online form of *LISA* began data accession somewhat later than its printed counterpart, it was decided to extend its coverage and to consult the printed forms of both of these services, from their inception through to the early 1980's; the index of each volume was searched for terms containing the appropriate word-stem.¹ The search of each database was initially terminated at the end of the accession year 1986, and the initial literature sample was selected from the analytical-level documents retrieved in each search. This seed sample was subsequently added to by extending the search of the databases forward to the end of the accession year 1989, and then to the end of the accession year 1992, though only documents published prior to 1990 were selected.

¹*LISA* became available online in 1976; its earliest accession year is 1969. The printed form of the service was titled *Library Science Abstracts* or *LSA* from 1950 to 1968 (Vol.1-Vol.19), and thereafter *LISA* (1969, Vol.20 --). The Name and Subject Indexes were used for *LSA* volumes, and the Subject Index for *LISA* volumes. *ISA* became available online on DIALOG in August 1983; its earliest accession year is 1966. The printed form of the service was titled *Documentation Abstracts* from 1966 to 1968 (Vol.1-Vol.3), and thereafter *Information Science Abstracts* (1969, Vol.4 --). The Subject Index was used. In both printed services, Cumulative Indexes were also consulted.

Entries in Relation to Accession Year. Figure 3-1 shows the total number of 'records' or 'entries' (document surrogates) retrieved from each of the three specialised databases by the end of the accession year 1992, plotted against accession year. Results from the print and online forms of *LISA* and *ISA* are plotted separately. Results from each Service are as follows:

- In *Library Literature*, 101 entries were indexed under the subject term 'BRADFORD'S LAW'; no other terms containing '-BRADFORD-' were appropriate. Entries first appear in 1972 and reach a maximum number of 12 entries per year in 1981; thereafter till 1992, approximately five documents have been appropriately indexed each year.²
- In *LISA online*, 115 entries had suitable descriptors; these will be detailed below, though it may be noted here that the predominant forms contain 'BRADFORD'S LAW'. A number of entries with inappropriate '-BRADFORD-'~containing descriptors were excluded. Entries first appear in the accession year of 1969, and reach a maximum of 13 entries per year in 1978; thereafter till 1992, approximately six documents have been appropriately indexed each year. In the printed form of the database -- which actually commences as *Library Science Abstracts* or *LSA* in 1950 -- the first suitable entry occurs in 1967, and there is no entry in 1968. (This extra 1967 document will be subsequently appended to the entries obtained online, for some uses). From 1969 to 1974, there are slight differences in the number of entries between printed and electronic forms, but thereafter, at least to 1980 -- after which year it was no longer analysed -- the documents entered in the printed service each year are identical to those of the electronic service.³

²(1) In the Subject Index, other '-BRADFORD-'~containing entries refer to libraries located in Bradford (England); these are detailed in Section 6 of this Chapter. In the Author Index, '-BRADFORD-'~containing entries refer to documents by and about 'S.C. Bradford', and documents by and about other 'Bradford's'; these are also discussed later. With respect to S.C. Bradford's publications, it is not apparent here that even Bradford's (1934a) paper, which is now credited as the initiator of the topic, is on the topic; though this could be ascertained from a later (1985) reprint in the database. (2) Available volumes of *Library Literature* include more than one accession year per volume prior to 1978. The first publications indexed as on the topic appear in this database in 1972. For the years 1972-1977, the available volumes each contain two accession years, posing a problem of how to divide the publications listed in each volume into their respective accession years. A division was made by carefully considering the publication month and place of publication of each item, but it must be noted that the data presented in the Figure for *Library Literature* in these years may be somewhat in error.

³(1) The excluded entries fall into three groups. Nineteen entries contained descriptors which refer to various libraries located in Bradford (England); these will be discussed in Section 6 of this Chapter. Five entries, from 1986 onward, and which contained 'BRADFORD'S LAW' in the descriptor field, were *Current Research Reports*, i.e. reports of research in progress, rather than final publications. One entry (accession no. 81-5677) is an exact duplicate of another entry (accession no. 80-168), except with regard to the descriptor field. (2) Of minor importance for the present study: five items found in the printed service were *not* found in the online service by searching for '-BRADFORD-' in the descriptor field. (One was found in *LISA online* with other descriptors). They should have been accessioned in 1973 and 1974. Four of these were analytical-level documents, suitable for inclusion in the seed literature sample. Two of the documents arrived in the seed literature via the other databases, but the remaining two documents did not.

- In *ISA online*, 87 entries had suitable descriptors; these will be detailed below, though it may be noted here that somewhat under half contain 'BRADFORD'S LAW'. Two entries with inappropriate '-BRADFORD-'~containing descriptors were excluded. Entries first appear in the accession year of 1967, and reach a maximum of nine entries per year in 1979; between 1980 and 1990, very few documents were appropriately indexed. In the printed form of the database -- which commences, as *Documentation Abstracts*, in the same year as the online service begins accession, that is 1966 -- the first suitable entry occurs in 1967. There are no differences of importance for the present study, between the entries per year in the printed and electronic forms, at least through to 1983, after which year the printed service was no longer analysed.⁴

From inspection of the plots in Figure 3-1, and more particularly of those for the two older services, *Library Literature* and *LISA*, it seems reasonable to conclude that, prior to c.1965, either publication of papers on the present topic was negligible, or the field was simply not recognised as such by the A&I Services. The latter situation seems rather unlikely if the former were not also the case.

Initial Literature Sample. The initial or seed sample of documents was formed from the entries retrieved from the three selected A&I Services, as described above, but to the end of the accession year 1986, rather than to the end of 1992. Thus, from *Library Literature*, 72 appropriate entries were retrieved; from *LISA online*, 85 appropriate entries were retrieved; and from *ISA online*, 68 appropriate entries were retrieved. The 85 entries from *LISA* were supplemented with an entry from the printed form of the Service for accession year 1967, that is prior to the first accession year of its electronic form. From each set of entries, those representing monographic-level documents such as dissertations or reports, or those referring to previous entries, were removed. Thus, from *Library Literature*, five entries were excluded, leaving 67 suitable entries; from *LISA*, three entries were excluded, leaving 83 suitable entries; and from *ISA online*, five

⁴(1) The two inappropriate '-BRADFORD-'~containing descriptors referred to the library of the University of Bradford, Bradford (England); they will be discussed in Section 6 of this Chapter. (2) After this study had begun, *ISA* changed its indexing practice by modifying its Descriptor field and adding an Identifier field -- a practice abandoned in 1993, when both fields were retroactively combined into a Descriptor field. Some labour was required to minimise disruptions introduced by these changes. It also seems to the author that accessioned documents strongly on the topic, which would have been likely assigned '-BRADFORD-'~containing terms in former and later years, were not so readily assigned these terms from the early to the late 1980's. A shift in indexing policy would explain the relative paucity of suitably-indexed documents in this period. It must be remarked that continual changes of practice by A&I Services, while they may be necessary experimentation for improving the service offered, makes analyses like that performed here quite difficult.

entries were excluded, leaving 63 suitable entries.⁵ In total, 226 entries from the three databases were inspected, and 137 items representing distinct analytical-level documents were selected.

The 137 itemised documents were obtained and inspected. With respect to the language of publication, 101 (73.7%) are in English, 14 (10.2%) in Russian, six (4.4%) in Portuguese, and three (2.2%) in German; 10 of the remaining documents were published in eight other languages, and three were published simultaneously in two languages.⁶ All but one of the documents occur in serials, and these were published in 16 different countries.⁷ Serials containing the most documents are *Journal of Documentation* (with 28 or 20% of the documents), published in the U.K.; *Journal of the American Society for Information Science* (13 or 9.4%), published in the U.S.A.; and *Nauchno-Tekhnicheskaya Informatsiya. Seriya 2*. (9 or 6.5%), published in the former U.S.S.R. Most (113 or 82.5%) of the documents are original research papers or conference papers and the remainder are letters (14 or 10%), notes, or biographical papers. Discounting a 1934 paper by S.C. Bradford (1934a) -- of which reprintings in 1977 and 1985 were accessioned here -- the year of publication of the documents ranges from 1967 to 1986, with a mean of 1976.4 near the mid-range, and a median of 1977.5. A more detailed inspection of this time profile, extended to 1989, is given on p.80 in the paragraph headed *Analytical-level Documents in relation to Publication Year*. To summarise, from the foregoing analysis, it would be reasonable to expect that the topic, while predominantly researched in English-speaking countries, is truly international, and further, that it is of relatively recent origin and is currently viable.⁸

These 137 documents were studied carefully as representatives of the topic. A conception was developed of the topic, and this influenced the acquisition and the classification of further documents.

⁵Of the five entries excluded from *Library Literature*, three represented monographic-level documents and two referred to previously-entered documents (but not as exact duplicates); the excluded entries for the other databases represented monographic-level documents.

⁶These languages are Spanish and Hungarian with two documents each, and, with one document each, Japanese, Slovak, French, Romanian, Dutch, and Ukrainian. Two documents were published in Romanian and English, and one in Russian and English.

⁷These countries are: U.K., Netherlands, France, Spain, the former East and West Germanies, the former Czechoslovakia, Hungary, Romania, the former U.S.S.R., India, Japan, Australia, Canada, U.S.A. and Brazil; additionally, some publications may be correctly described as international. In this study, countries are taken to be those in existence over the greater part of the time interval -- 1960 to 1989, i.e. no adjustments have been made for the recent changes in the U.S.S.R., Germany, Czechoslovakia and Yugoslavia.

⁸It should be restated that the three A&I Services used here are produced in the U.K. (*LISA*) and the U.S. (*Library Literature* and *ISA*), but they do strive to provide an international coverage of the main serial publications in the discipline.

Subsequent Additions to the Seed Literature Sample. In due course, this sample of documents was expanded by later searches from the three A&I Services in the manner described, first to the end of accession year 1989, and subsequently to the end of accession year 1992. Only documents published prior to 1990 were included in the sample. The expanded later sample no longer served as a literature 'seed', but the new accessions in the three A&I Services provided additional checks on the accuracy of the topic selection criteria. From the first additional search, 16 documents were retrieved, and from the second additional search, ten documents (published before 1990) were retrieved; all were analytical-level documents.⁹ Thus, the literature sample from the three A&I Services finally included 163 documents.

Subsequent Re-evaluation of the Seed Literature Sample. Though the sample of literature from the three A&I Services provided a guide to the definition of the topic, this definition was formed gradually from a study of all documents acquired. Consequently, the seed literature, being a part of the wider collection, was continually re-evaluated and re-classified. Nevertheless, as the representative of the conventional view of the topic, the seed literature, or the predominant part of it, must remain in the collection of documents (most probably) on the topic, and any rejection of seed documents must be carefully explained. When the graphical selection criterion was finally stabilised, it admitted 148 (90.8%) of the 163 seed documents. Of the 15 rejected documents, seven -- including (reprintings of) S.C. Bradford (1934a) mentioned earlier -- could be re-introduced in special supplementary collections; the distribution of these documents is examined in Chapter 5, §5, p.404. Of the remaining eight documents, only one is judged to be a regrettable though not serious omission, while five others are judged to have no claim to inclusion at all. The exclusion decisions, which effectively repudiate judgements of the A&I Service experts, have been justified in detail in Section 4C, p.181.

Analytical-level Documents in relation to Publication Year. In Figure 3-2 (top), the number of analytical-level documents on the topic, as defined, that were 'correctly' indexed by each of the three A&I Services through to the end of 1992, are plotted against the year of publication of the document. Documents are given their first year of publication rather than their reprinting dates, so that the paper of S.C. Bradford (1934a), if included in the collection, would be plotted correctly at 1934; excluding this paper, no other document appears until 1966. The three plots are in moderate agreement, except

⁹The subsequently added seed documents are not notably dissimilar (except in publication year) from the earlier seed documents. Of the 16 documents added by searching to the end of 1989, 12 documents were published in English, three in Japanese, and one in Russian; publication years ranged from 1986 to 1989, though one document was actually a reprint of a 1971 document. Of the ten pre-1990 documents found by searching to the end of 1992, seven documents were published in English, two in Russian, and one in Portuguese; publication years ranged from 1986 to 1989.

for the initial delay in *Library Literature* accessioning, or appropriately indexing, documents on the topic. All show a period of high publication from 1977 to 1981, and two show a minor burst of publication around 1969, shortly after the probable inception of the topic. The combined data from the three A&I Services are plotted in Figure 3-2 (bottom).¹⁰ The initial 'burst' of documents in 1969, and the later period of high production are quite obvious. This plot provides a guide to the time-course of the topic literature.

Terms used by A&I Services to index Topic. As well as providing a seed literature sample and an estimate of the time-course of the topic, A&I Services were consulted as to the form of the terms used to index documents on the topic. It was hoped that this would provide a sample of characteristic or important words used in the topic to assist in the analysis of the seed literature. In addition to the three specialist A&I Services previously mentioned, three A&I Services which treat Library and Information Science within a broader subject coverage were consulted. These are: *ERIC* (1966-), an educational resource database; *INSPEC* (1969-), a database for Physics, Electronics and Computing; and *PASCAL* (1973-), a multidisciplinary database. These additional databases were searched online through the DIALOG system for the occurrence of the word-stem '-BRADFORD-' in the Descriptor or Identifier fields, that is in fields which are provided by the Service specialists. Only those entries with '-BRADFORD-'-containing index terms which were judged appropriate to the topic were selected. Entries were not further judged according to the present author's definition of the topic. The results reported here are for final searches made to the end of the accession year 1992, and include both monographic-and analytical-level documents published prior to and in that year.

A summary of terms (descriptors, identifiers, subject headings) used by the six A&I Services to index documents on the topic is provided in Table 3-1. The exact terms are listed in the left column, in various groupings. Numeric values in the table are the percentages of all suitable documents in each database which contain each term, exactly as printed, in the appropriate field. Thus, for example, it can be seen that for *ISA*, 44.2% of the 87 suitably-indexed documents contain the term 'BRADFORD'S LAW' in their descriptor field. In some cases, the full index term may be longer than the exact term shown; for

¹⁰(1) There are 154 distinct documents in this plot. As noted in the previous paragraph of the main text, 148 are retained in the (main) topic collection defined by the (final) graphical selection criterion, while six -- after the exclusion of Bradford (1934a) because of its early publication date -- are placed in the small supplementary collections. Eight of the 163 seed documents were taken to be 'incorrectly indexed'. (2) Two documents in *ISA* required corrections to their publication dates: one (accession no.7301526) from '1672' to '1972', and one (accession no.9201292 {online}, 9200840 {printed}) from '1991' to '1988'.

example 'BRADFORD'S LAW' might be part of the full descriptor 'BRADFORD'S LAW USE IN PHYSICS'. Further, in some cases, more than one descriptor (or etc.) containing a suitable form of '-BRADFORD-' may be given to a document; fractional rather than whole occurrences were then assigned to each such descriptor (or etc.) in that document. In the example above, 44.2% of 87 documents is $38.5 = 35 + 7(0.5)$ documents; that is in *ISA*, 35 documents have as their '-BRADFORD-'-containing descriptor(s) only the form 'BRADFORD'S LAW', but 7 documents also contain another form, for example 'BRADFORD'S DISTRIBUTION', and are only assigned a half-occurrence each.¹¹ With respect to each A&I Service, we may note:

- In *Library Literature*, all 101 publications retrieved were placed under the subject heading of 'BRADFORD'S LAW'.
- In *LISA online*, 87.4% of the 115 publications contain 'BRADFORD'S LAW OF SCATTER' in the Descriptor field. An alternative term, 'BRADFORD-ZIPF DISTRIBUTION', was favoured only in 1969-1970. The same pattern is shown in the index headings of the printed form of *LISA*, but here additional alternative terms were used; two, not already listed in the Table from other databases, are 'BRADFORD, S.C. LAW OF SCATTERING' (in 1967) and 'MANDELBROT-ZIPF-BRADFORD DISTRIBUTION' (in 1978).¹²
- In *ISA online*, 44.2% of the 87 publications contain 'BRADFORD'S LAW' in the Descriptor field, though with slight variations included, this value increases to 60.3%. Between c.1973 and c.1980, the alternative forms 'BRADFORD- DISTRIBUTION' were favoured. A

¹¹Correctly, the assignment of occurrences is a little more complex. For example, permutations of one descriptor were treated as replicates and ignored; with a descriptor field such as: 'BRADFORD'S LAW OF SCATTER; SCATTER LAW, BRADFORD'S; LAW OF SCATTER, BRADFORD'S; BRADFORD DISTRIBUTION', the first three descriptors are together assigned an occurrence of 0.5 only, while the fourth descriptor is also assigned an occurrence of 0.5. Therefore, in no cases was it necessary to assign fractions smaller than one half.

¹²Comments are needed on these results. (1) From c.1974 to c.1977 especially, the number and length of descriptors increased substantially in this database, each descriptor containing a number of separate components which might be permuted, internally and externally, with additions and deletions, to obtain other descriptors. Thus, as mentioned, the term listed in the Table is often contained in longer descriptors, it may be replicated, and it may occur in permuted forms, e.g.

'BRADFORD'S LAW OF SCATTER (CITATION ANALYSIS: PERIODICAL ARTICLES: TOBACCO AND HEALTH); SCATTER LAW, BRADFORD'S (CITATION ANALYSIS: BIBLIOMETRICS: PERIODICAL ARTICLES: TOBACCO AND HEALTH); SMOKING AND HEALTH BULLETIN (NATIONAL CLEARINGHOUSE ... (USA) STUDY: BRADFORD'S LAW OF SCATTER: CITATION ANALYSIS: BIBLIOMETRICS: PERIODICAL ARTICLES: TOBACCO AND HEALTH); '.

In such a case, only one phrase would be counted, viz. 'BRADFORD'S LAW OF SCATTER'. (2) At the end of accession year c. 1992, seven records already entered could only be retrieved online using a new *Note* field. It was decided to retain them. Correctly, then, three additional documents in this new *Note* field -- which were never located through their Descriptors per se -- should also be included. As noted elsewhere, changes of practice by A&I Services, while presumably necessary for their main role, make analyses like that performed here difficult. (3) The term 'SCATTER' used alone, in either electronic or print form, while frequently retrieving topic documents, also retrieves what I judge to be unrelated, or only distantly related, documents.

variety of terms are used in this database and most occur only in one document each. The same pattern is shown in the index headings of the printed form of *ISA*.¹³

- In *INSPEC online*, 32.6% of the 46 suitable publications have 'BRADFORD'S LAW' as an Identifier. A considerable variety of terms are used, most only in one document each.¹⁴
- In *ERIC online*, 95.8% of the 48 suitable publications have 'BRADFORD LAW', and 91.7% of suitable publications have 'BRADFORD LAW OF SCATTER', as an Identifier. There is very little variation in the form of the indexing terms used.¹⁵
- In *PASCAL online*, 96.6% of the 58 suitable publications have 'BRADFORD LAW' as a Descriptor. There is very little variation in the form of the indexing terms used.¹⁶

In summary, 29 different forms of '-BRADFORD-~containing terms were used by the six A&I Services to index the documents their specialists judged to be on the topic. A crude estimate of the likelihood of such a document being assigned each term in the combined databases is provided in the right margin of the Table: the estimate is the mean of the six

¹³ Longer descriptors were favoured in this database initially, so as mentioned, the term listed in the Table is often contained *in* a descriptor, though rarely more than once per document. Complications introduced by the introduction of an Identifier field in the mid-1980's and its removal in c.1993 have already been noted.

¹⁴ 101 entries were retrieved from *INSPEC online* by searching the Identifier field for the word-stem '-BRADFORD-', from its first accession year of 1969 to the end of the accession year 1992. 46 entries were appropriate to the topic; most of the remainder had some connection to Bradford (England); and see Section 6. The first suitable entry appeared in accession year 1971, with c. 2 entries being accessioned each year subsequently. The printed form of the A&I Service was also checked without uncovering any additional relevant information; the volumes consulted were: *Control Abstracts*. Vol.1-Vol.3 (1966-1968), and *Computer and Control Abstracts*. Vol.4-Vol.15 (1969-1980).

¹⁵ 80 entries were retrieved from *ERIC online* by searching the Identifier field for the word-stem '-BRADFORD-', from its first accession year of 1969 -- subsequently extended back to 1966 -- to the end of the accession year 1992. 48 entries were appropriate to the topic; most of the remainder referred to educational institutions, e.g. in Bradford (England); and see Section 6. The first suitable entry appeared in accession year 1969, with c.2-3 entries being accessioned each year subsequently. The printed form of the A&I Service was also checked without uncovering any additional relevant information; the volumes consulted were: *RIE (Resources in Education)*.1967-1978, and *CIJE (Current Index to Journals in Education)* Annual and Semi-annual Cumulations 1969-1980.

¹⁶ 157 entries were retrieved from *PASCAL online* by searching the Descriptor field for the word-stem '-BRADFORD-', from its first accession year of 1973 to the end of the accession year 1992. Only 58 (37%) entries were appropriate to the topic; the remainder referred to documents on a wide variety of themes, e.g. a biochemical method, geological deposits, and a disaster at Bradford (England); and see Section 6. The first suitable entry appeared in accession year 1975, but few entries occur till after 1983. No distinction has been made between English, French and Spanish Descriptor fields, with the English equivalent form, even if not specifically given, being used here. Until c. 1990, the '-BRADFORD-~containing descriptor occurred primarily in French only; 25 of these ('LOI BRADFORD', 'LOI DE BRADFORD', and 'DISTRIBUTION BRADFORD-ZIPF') have been translated into English. Of the 32 English descriptors provided simultaneously with the French descriptors, one is misspelled ('BRADFORT LAW') and has been corrected here. The printed form of the A&I Service, *Bulletin Signalétique. Section 101: Information Scientifique et Technique*. was checked from 1970 to 1973 (Vol.31-Vol.34); one new item was uncovered in accession year 1972 under an (English) subheading 'STATISTICAL ANALYSIS, BRADFORD'S LAW', but this has not been included here.

percentages in each row, where blank cells in the row represent zero values. The most frequently-used terms, unsurprisingly, are 'BRADFORD'S LAW' and 'BRADFORD LAW', with 'BRADFORD'S DISTRIBUTION' and 'BRADFORD DISTRIBUTION' the next most used. If this distribution reflects that of different words adjacent to occurrences of 'BRADFORD' in the text of documents on the topic, then the prospect appears favourable for characterising, and possibly diagnosing, most documents with a small number of phrases. That is, the prospect of obtaining a succinct graphical selection criterion for the topic appears to be good. However, the great variety of rarely used terms suggests that the characterisation of all documents, and the construction of a selection criterion with perfect recall, may be more difficult. The use of the descriptors of A&I Services to provide a good graphical selection criterion for a topic will be considered further in Chapter 4, Section 6, p.378.

Re-examination of the Role of A&I Services in the Present Procedure. The above analysis of the terms used to index the topic by A&I Services may be considered from a different perspective, that of the soundness of the general procedure for forming a collection. The procedure rests on the assumption that 'research topic' is the minimum taxonomic category for research knowledge recognised by specialist A&I Services, and is identified as the narrowest fixed 'descriptor' of the A&I Services' thesauri. The above analysis shows that of the three specialist A&I Services used to form the collection, only *Library Literature* conforms to the model, with 100% of its 'topic' documents indexed under one term, 'BRADFORD'S LAW'. *LISA* conforms to a lesser degree, with 87% of its 'topic' descriptors containing 'BRADFORD'S LAW OF SCATTER', albeit possibly permuted. *ISA* conforms least, with 60% of its 'topic' descriptors containing 'BRADFORD'S LAW' or 'BRADFORD LAW', albeit often in long strings. The conclusion to be drawn is that greater care must be given to select only those A&I Services which use terms from a controlled vocabulary, and so conform to the classification model, and to reject those A&I Services which effectively use natural language descriptors, and so do not. However, it might be argued that the problem is one of my own making -- that afterall, 100 documents in *LISA* and 52 documents in *ISA* do at least contain the term 'BRADFORD(S) LAW' in their descriptors, and that, by using the stem '-BRADFORD-', even by my own model, I have incorrectly conflated distinct terms and conflated separate 'topics'. (According to the model, as *ISA* views the situation, for example, 'BRADFORD DISTRIBUTION' was, for a time, a distinct topic from 'BRADFORD'S LAW'). I have used my own preliminary interpretations of some 'topic' to sift through these A&I Services' interpretations of topics, though the latter are supposed to be the sole, or at least the standard, interpreters. Though unlikely to be a contentious choice, this has introduced an arbitrariness into the procedure at the

very outset. Correctly, I should have selected from each Service just one suitable descriptor.¹⁷

¹⁷This 'single descriptor' might include grammatical permutations, and, more questionably, embedding in longer strings. Choosing only a single descriptor could create further problems for the extraction of a diagnostic pattern against the A&I Services' near background samples. The reality is that A&I Services provide only an approximate guide to a consensus of interpretations on topics amongst researchers, and my own initial understanding may be a better estimation of this consensus. But there is no immediate way of determining this, which is the reason for using the A&I Services as a standard interpreter.