Mapping the knowledge base of information policy: clusters of documents, people and ideas

Volume 1: Argument

Ian Rowlands
Information Policy Unit
Department of Information Science
City University

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Information Science)

April 1998

External examiners

Professor Mairéad Browne

Dean, University Graduate School, University of Technology Sydney

Professor Charles Oppenheim

Professor of Information Science, Department of Information and Library Studies, University of Loughborough

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me.

This permission concerns only single copies made for study purposes.

Acknowledgements

A number of people made a valued contribution during the course of this work. *Primus inter pares*, I would like to thank my advisor, David Bawden, for his constant support, advice and keen insights—I am especially indebted to him for his guidance on the choice of appropriate multivariate techniques.

I am grateful to Robert Kimberley, Managing Director of the Institute for Scientific Information's European Branch, who granted me permission to use the company's citation indexes for the purposes of this research.

Stephen Robertson played a key role—usually at points of grave personal confusion—and I thank him for his patient advice on some tricky technical issues Micheline Beaulieu gave generously of her time to read and comment on the final draft. Peter Williams prepared the data for the author cocitation analysis as well as playing detective, locating a rich supply of background information on the various authors included in the cocitation study. Tamara Eisenschitz and Paul Turner maintained subtle pressure throughout by reminding me of the need to complete this thesis. Did they collaborate, I wonder?

If Tamara and Paul provided the stick, Dave Nicholas offered a juicy carrot in the form of his generous and encouraging comments on various early drafts. I would also very much like to thank Mairéad Browne of the University of Technology, Sydney, for her constructive criticism of my initial project ideas.

Contents

PART TWO: EXPERIMENTAL EVIDENCE

List of tables, figures and coding frames

Tables

Figures

Figures *continued*

Content analysis coding frames

Abstract

This thesis investigates aspects of the intellectual and social structure of the field of information policy through a detailed examination of the serials literature. The aims of the research are to explore how information policy scholarship is organised—in terms of its relation to other fields and disciplines; whether it constitutes a distinct specialty in its own right; and what kinds of institutional structures and arrangements exist to support research and scholarship.

In Part One, a literature review identifies previous bibliometric and other studies which are relevant to studies of scholarly disciplines and knowledge communities. It discusses the interdisciplinary problem-oriented nature of information policy and considers some of the modes of enquiry which characterise investigations this area.

Part Two consists of a series of experiments carried out on a test collection of 771 periodical articles drawn from the *Social Science Citation Index*. The empirical work comprised four linked studies: a bibliometric census study; an analysis of document clustering; an author cocitation study; and a content analysis. Extensive use was made of multivariate statistical techniques, notably principal components analysis, hierarchical clustering, discriminant and correspondence analysis to identify statistically significant and meaningful patterns and structures within the test collection.

The study concludes that information policy is a growing and reasonably distinctive field of study with strong links to library and information science, law, media studies, and the political sciences. It is suggested that the field is not unified and that research is still primarily organised along national and traditional disciplinary lines, with little evidence of significant collaborative activity across institutions or sectors. The research base is highly dispersed, with practitioners playing a major role in the production of knowledge. In institutional terms, the field is very thinly spread, with few signs of concentration.

Overview

Chapter 1: Overview

> *"Any group shares an overall sense of orientation or purpose its charter in anthropological terms"*[1]

1.1 Motivations behind the research

This thesis examines the intellectual and social structure of an area of scholarship identified by its focus on the policy issues associated with information and its use in society. This topic is currently attracting much attention, and the term 'information policy' has gained wide currency within and beyond the library and information science community[2]. As well as a growing archive of literature on the topic, it is now possible to identify a community of information policy analysts and scholars employed in academia, in think tanks, and in government. To some commentators, information policy represents an emerging field of study in its own right (Braman, 1990; Burger, 1993).

Information policy is an example of what a political scientist might refer to as an 'issue area', one which groups together concerns relating to the same subject: housing or social justice, for example. Information seems only recently to have become recognised as an issue area in its own right (Braman, 1990) and, as such, it suggests some particularly interesting questions in relation to how new areas of policy study emerge and become established. This thesis addresses some fundamental questions about the nature of information policy scholarship. Does information policy constitute a field of study in its own right? Where does it stand in relation to neighbouring disciplines? What is the intellectual and social structure of information policy research? What factors characterise and define the information policy research community? The subject of the thesis is therefore the *field* of information policy, not information policy as such.

---

[1] Rennie (1977 221).

[2] The term 'information policy' occurs in 296 records in DIALOG's *Social SciSearch* in the period to 31 July 1997. 79 of these mentions fall outside ISI's library and information science journal category

The consensus of those few authors who have written about information policy as a distinctive field of studies is that it is at a very early stage of intellectual development, with little agreement on what precisely the field comprises. Kajberg & Kristiansson state the problem very clearly:

> *"A detailed examination of the published record of information policy reveals a variety of definitive and analytical approaches to the concept of information policy. There is an evident lack of consensus of what constitutes the core of information policy. An array of classificatory approaches in the field are offered but what is missing is a coherent theoretical framework"* (Kajberg & Kristiansson, 1996:5).

This view is endorsed by Aldhouse, a data protection specialist, who finds that:

> *" . as a stranger to the discourse of information policy, I have found some difficulty in identifying a unified set of topics which might be the subject of something called 'information policy'. Indeed if one considers the nine categories referred to by Rowlands (1996), the stranger might readily conclude that the only element unifying information technology policy, intellectual property, information disclosure, confidentiality and privacy, and the others is that they are all of concern to librarians and information scientists"* (Aldhouse, 1997:115).

Browne, a leading Australian information policy researcher, admits with refreshing honesty that.

> *"... there is a feeling of discomfiture in a field in which its researchers are unable to even broadly indicate the substance of what it is they desire to study"* (Browne, 1996).

If information policy is built on such weak epistemological foundations, is it proper to speak of it as a field of studies in its own right? To begin to answer this question, we need to reflect on what criteria might be used to construct a working definition of a 'field of studies'. King & Brownell (1966) argue that such

a definition should embrace a wide range of factors, not simply reference to some objective system of knowledge. In their account, fields of study are identifiable by a shared sense of community and belonging, a network of communications, a tradition, a particular set of values and beliefs, a domain, a mode of enquiry. In other words, fields of study may be best understood not by succinct definitive statements, but by what people actually *do*.

Thus, one approach to understanding the nature of information policy as a scholarly activity might lie in an exposition of its communication patterns using bibliometric techniques. Probably the most widely cited definition of bibliometrics is that of Pritchard. In his view, the aims of bibliometrics are:

> *"to shed light on the processes of written communication and of the nature and course of development of a discipline (in so far as this is displayed through written communication), by means of counting and analysing the various facets of written communication"* (Pritchard, 1968, cited in Borgman, 1990:13).

While acknowledging that bibliometric methods can only shed light on the formal surface aspects of scholarly communication, they do offer the possibility of insights which are unique, if inevitably limited. Borgman (1990) notes that bibliometric experiments have frequently been used to respond to research questions such as: "What is the scholarly community of X?" or "Of what types of scholars is the community composed?" and she cites a number of highly influential studies and authors[3] in this area.

Borgman notes that studies of this kind raise an important theoretical issue, in the sense that we are here combining studies of 'invisible colleges' and studies of scientific specialties, which, although theoretically distinct, have much in common methodologically. The links between bibliometrics, scholarly communication and the research aims and objectives which follow are more fully developed in Chapter 2.

---

[3] Including: Leydesdorff, 1989, Lievrouw, 1990, Pierce, 1990; Price, 1965; Small, 1973; White & Griffith, 1981

1.2 Research aims

The broad aims of this thesis are illustrated as a series of research questions in Figure 1.1, organised under four related headings. If information policy is a reasonably coherent domain, one would expect, for example, to be able to find empirical evidence to show that there are common paradigms and frameworks; journals where research findings are regularly reported; and networks of scholarly activity which transcend traditional disciplinary boundaries.

Figure 1.1: Overarching research questions

Disciplinary boundaries

To what extent is 'information policy' a distinct specialty in its own right, or does it comprise several specialist areas which are primarily dependent on other disciplines (e.g information science, policy studies, law)?

Invisible colleges

What is the social and collaborative structure of information policy research? Are there identifiable 'schools' or networks of researchers? If so, how do these schools map onto the disciplinary mapping above?

Institutionalisation

What kind of institutional arrangements (e.g. research centres, collaborative studies, specialist journals) currently support information policy research? Are there signs that knowledge production is becoming more concentrated?

Knowledge production

How should information policy research be characterised in relation to the sociological model of knowledge production developed by Gibbons and others (1994)?

These research above are addressed by developing a series of experimental bibliometric indicators. These indicators are contextualised as far as possible and linked to external sources of evidence from the literature and (in Chapter 8) a validation questionnaire.

Subsidiary research aims

In addition to these large-scale research questions, the thesis re-examines some existing conceptual models in the information policy literature. For example, in a 1986 paper, Trauth argued that there was a significant gap in the information policy literature for studies that combined a high degree of interdisciplinarity with a focus on the immediate needs of policy-makers. Trauth illustrated her argument by example rather than systematic observation, and so a subsidiary research aim is (a) to empirically validate Trauth's claims made in the mid-1980s, and (b) to review the situation ten years on.

Similarly, there are other descriptive models in the information policy literature which have not yet been subject to systematic investigation. These include:

- Kajberg & Kristiansson's Information policy scale model (1996)
- Rowlands' Typology of information policy methodologies (1996)

1.3 The research strategy

The structure of the thesis is in three parts:

- Part One an extended literature review
- Part Two: a series of empirical studies
- Part Three: a synthesis of the findings

A summary of each of the remaining Chapters and the specific research objectives follows.

Part One: Evidence from the literature

The study of specialties (Chapter 2)

This first section of the literature review reflects on the fundamental units of analysis with which bibliometric studies are concerned: documents, people and ideas. It identifies relevant antecedants for the present study (drawn from work in bibliometrics, scholarly communication and the sociology of knowledge production) and provides a rationale for the experimental work reported in Part Two.

The field of information policy (Chapter 3)

This section briefly reviews that literature which deals, at a rather abstract level, with the discourse of information policy. It focuses on the interdisciplinary, problem-oriented, nature of information policy, its normative structure, and some of the characteristics which differentiate 'information' from other public policy issues.

Information policy research methods (Chapter 4)

The final section of the literature review considers some of the methodologies and approaches to understanding which previous authors have brought to bear in carrying out information policy studies. The material in this section introduces a number of frameworks which are later used in the content analysis phase of the experimental design (Chapter 9).

These three introductory chapters are integral to the research methodology in that they are driven by the need to identify concepts and frameworks which can be used to develop useful indicators of the structure of the information policy serials literature.

Figure 1.2: Literature review: research objectives

To review pertinent literature in three subject domains: bibliometrics, scholarly communication, and the sociology of knowledge production. This material justifies the experimental approach adopted in Part Two (Chapter 2)

To review existing writings on information policy as a discourse (Chapter 3).

To provide working definitions of some basic terms and concepts used in information policy (Chapters 3, 4)

To identify useful definitive and analytic frameworks from the policy studies literature for subsequent content analysis (Chapter 4)

Part Two: Experimental evidence

The research design (Chapter 5)

The next part of the thesis reports on an empirical investigation into the structure of the information policy serials literature. This comprised four studies: a bibliometric census study; a cluster analysis; an author cocitation study, and a content analysis. As well as being implicitly linked by reference to common research aims, each of the experiments shared a common data platform: a document test collection specifically created for the purpose of this research. The rationale behind each of the experiments is outlined below.

Simple bibliometric analysis (Chapter 6)

Bibliometrics is concerned with the quantitative study of literatures as reflected in bibliographies. Simple bibliometric indicators such as the number of authors active in a field, or the productivity of different nations or types of institutions may be valuable in offering a broad overview of a field of scholarly activity. Indeed, some writers have gone so far as to use bibliometric indicators to evaluate aspects of science policy (see, for example, the review by White & McCain, 1989).

This Chapter examines the distribution of various bibliometric elements across the test collection and through time. The approach is rooted in Borgman's (1990) framework which argues that bibliometric studies deal essentially with the distribution of three kinds of variables: *documents* (e.g. Bradford core and scatter, median age of citations), *people* (e.g. numbers of authors per paper, corporate addresses) and *concepts* (e.g. journal categories, subject headings).

Figure 1.3: Simple bibliometric analysis: research objectives

To develop a quantitative understanding of the following communication processes.

- *growth in the information policy serials literature*
- *concentration and dispersal of articles across journal titles (Bradford core and scatter)*
- *knowledge accumulation (Price's index)*
- *ageing and obsolescence*
- *patterns of authorship and author productivity*
- *collaboration between authors and between institutions*
- *institutional production of articles*

Cluster analysis (Chapter 7)

This Chapter employs hierarchical clustering techniques to investigate the presence of any convincing underlying natural structure in the document test collection, based on classifications generated from a range of bibliometric and content based indicators. The Chapter addresses the question of whether the bibliography is best considered as a single homogenous entity, or as a series of joint bibliographies distinguishable by some, initially unknown, criterion or criteria

Figure 1.4: Cluster analysis: research objectives

> *To explore the structure of the test collection using automatic classification techniques, in order to determine whether the test collection is best regarded as*
>
> *• a single homogenous bibliography, or*
>
> *• a series of joint bibliographies*

Structural bibliometric analysis (Chapter 8)

Author cocitation analysis is a methodology which, it has been claimed, offers unique insights into the social and intellectual structure of a field of study. This Chapter explores the patterning evident in the cocitation patterns of 21 leading information policy authors using a battery of multivariate statistical techniques. In author cocitation studies, the unit of analysis moves up from individual articles (as in Chapters 6,7 and 9) to the level of an individual author's writings or oeuvre. By analysing patterns of cocitation, it may be possible to identify authors who are central or peripheral to a field, and more locally, authors who are central or peripheral within specialities. The technique therefore offers a potentially useful set of data for mapping disciplines. It provides an example of what Marshakova-Shaikevich (1993) terms 'structural' bibliometrics; unlike the approach in Chapter 6 the emphasis here is on the connections between objects rather than on 'simple' distributions. The findings are corroborated by means of a simple postal questionnaire to the authors included in the study.

Figure 1.5: Author cocitation analysis: research objectives

> *To identify key authors within the information policy community's 'intellectual frame of reference'*
>
> *To identify social and collaborative networks of authors*
>
> *To identify relationships between different topics within information policy*
>
> *To identify relationships between information policy and neighbouring disciplines*

Content analysis (Chapter 9)

The simple bibliometric analysis presented in Chapter 6 was based on elements which were either already present in the records downloaded from the *Social Science Citation Index*, or which could be derived with little effort. The content analysis extends this earlier analysis by investigating the distribution of various *content* indicators inferred from the original full-text documents. These indicators are qualitative and they codify and describe such factors as the aims, scope and intentions of published articles; the methodologies used; and the stage of the policy cycle under investigation. The intention was to see whether any patterning in the distribution of these elements was evident across topics, geography and through time.

Figure 1.6: Content analysis: research objective

> ***To extend the simple bibliometric analysis in Chapter 6 by examining the distribution of various qualitative variables across the test collection.***

Part Three: Synthesis of findings

Conclusions (Chapter 10)

The concluding Chapter attempts to bring together the findings of the four experimental studies. It acknowledges the provisional, incomplete, nature of the evidence and suggests a number of areas which require further investigation.

Part One
Evidence from the literature

Chapter 2: The study of specialties

> *"There are very strong social organisations underlying scientific work, possibly some of the strongest voluntary organisations ever studied"*[4]

2.1 Bibliometrics and the study of specialties

A variety of approaches to the analysis of scientific knowledge have been adopted over the past thirty years, ranging from the philosophical (Popper, 1959; Mullins, 1968) to the sociological (Kuhn, 1962; Crane, 1972; Merton, 1973)[5]. In the postscript to the second edition (1970) of his classic work, *The Structure of Scientific Revolutions*, Kuhn argued that much more empirical work was needed to support his ideas about the community structure of science and cited work by a number of information scientists and sociologists (Garfield, 1964; Hagstrom, 1968, Price, 1965) who were beginning to establish bibliometrics as a powerful research tool, one capable of throwing new light on the nature of scientific communication.

This thesis offers an account of the intellectual and social characteristics of the information policy specialty, in so far as this is possible from an examination of the serials literature. The research design is highly quantitative and bibliometric techniques are used to identify clusters of related documents, people and ideas The underlying assumption is that the patterns that may emerge from such analysis carry meaning and that this meaning may be interpretable. The present Chapter acknowledges the strengths and the limitations of the bibliometric approach and tries to position these documentary techniques in a broader intellectual context. It examines the bibliometric literature in an attempt to identify useful tools and frameworks for the remaining sections of the thesis

---

[4] Griffith (1990:43).

[5] An useful discussion of the influence of some of these thinkers, notably Merton, on bibliometric theory can be found in Pierce (1990).

2.1.1 The development of bibliometrics as a research tool

The term bibliometrics refers to a battery of quantitative techniques which are used to organise data about communication artifacts—typically research articles published in scholarly journals—and to represent those data in a meaningful way (Lievrouw, 1988). Bibliometrics is a major area of activity within information science and is increasingly receiving attention from outside that field, notably in communication research (see, for example, Lievrouw, 1988 and 1990, Borgman & Rice, 1992) and in social and philosophical studies of science (Marshakova-Shaikevich, 1993).

The scope of bibliometric inquiry is wide, ranging from Bradford studies of documentary core and scatter phenomena, studies of author productivity, citation analysis and work on the ageing of literatures, to the development of science policy indicators (White & McCain, 1989). Regardless of the research questions asked, however, Marshakova-Shaikevich argues that two distinct modes of bibli metric inquiry have emerged over the past thirty years:

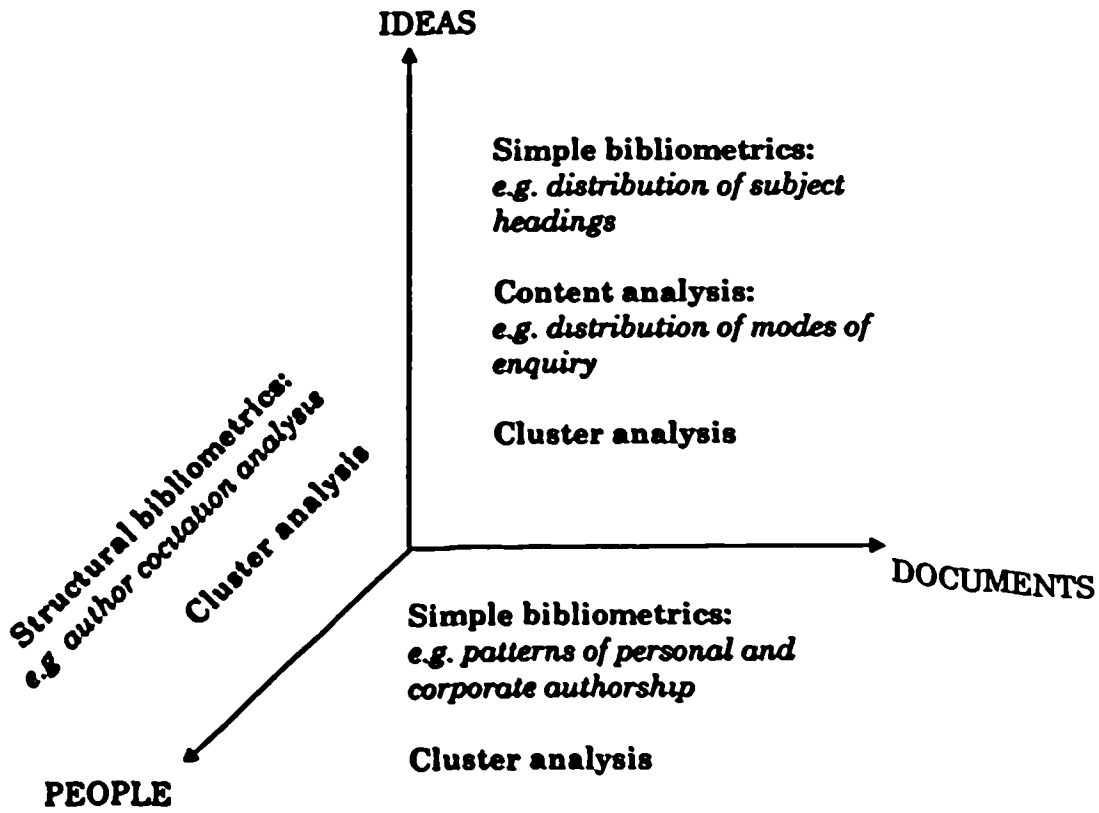
> *"The first consists in following the dynamics of separate objects: publications, authors, keywords from publication titles, distributions by countries or subject headings in journals, etc. The second consists in identification of links between objects, their correlation and classification. Bibliometric studies of science related to the first approach may be named 'simple' bibliometrics', to the second—structural bibliometrics"* (Marshakova-Shaikevich, 1993 5).

The emergence of the structural perspective in bibliometrics is of particular significance in the context of studies of scholarly communication and is probably the main reason why there has been a resurgence of interest in bibliographic tools among communications researchers. As the amount of data available from the scholarly record has increased, so more and more sophisticated bibliometric tools have been developed, often in conjunction with powerful statistical clustering techniques borrowed from other social sciences. Bibliometric data are particularly useful for studying trends in scholarship because of the massive

datasets that can now be obtained in machine-readable formats; indeed, some authors claim that virtually no other conventional method (surveys or case studies, for example) can provide anything like as comprehensive a coverage of a specialty or disciplinary field. Bibliometric techniques offer the additional benefit that they are unobtrusive, reasonably reliable, and may be replicated by others (Pierce, 1990).

While published bibliometric studies evidence a wide range of techniques and theoretical approaches, it is possible to argue that in all cases only three fundamental variables are examined: *producers* of communication, communication *artifacts*, and communication *concepts* (Leydesdorff, 1989; Borgman, 1990). This is a powerful concept and one which is expressed in the subtitle of this thesis as 'clusters of people, documents and ideas'.

Figure 2.1: Clusters of documents, people and ideas[6]



[6] Adapted from Leydesdorff (1989).

The concept is represented in Figure 2.1 on the previous page. It also locates the principal research methods used in this thesis in relation to the three variable axes. Leydesdorff (1989) points out that each variable may be studied at various levels of aggregation. Hence, 'people' may be studied as individuals (for example, by determining an author's reputation through an analysis of his/her citations), or as members of research groups or even national research communities. Similarly, 'documents' may be analysed at the level of individual articles (as in citation 'superclassic' studies) or in aggregated form as journals or author *oeuvres* (as in author cocitation studies), or whole literatures (as in bibliometric census studies). 'Ideas' may similarly be studied at various levels, including theories, specialties and disciplines.

The point of the model is that it provides an organising framework for data collection and analysis, and a reminder of the need to distinguish clearly between units of analysis and levels of aggregation. It also offers a framework potentially linking bibliometric studies with work in the philosophy and sociology of science (lying principally on the ideas-people plane). Leydesdorff (1989) argues that studies of specialties should place 'specialty' in the perspective of either social structure, cognitive structure, or scientific communication. While arguing that studies of specialties should ideally draw upon more direct (i.e. behavioural) data than can be obtained from a secondary examination of the published record alone, Leydesdorff acknowledges that bibliometric data is somewhat richer in these resonances than is sometimes assumed.

> *"Although the bibliometric tradition deliberately confines its domain to texts, it may, however, reveal regularities and patterns in scientific communication which are not consciously available to the actors involved—and therefore should not be asked of them—but yet structure their behaviour"* (Leydesdorff, 1989:338).

Happily, the bibliometric literature contains many examples of work which bear a scientific communication orientation. Borgman (1990) argues that bibliometric

techniques have been used to tackle four generic research questions in scholarly communication:

- to characterise scholarly communities (e.g. Small, 1973; White & Griffith, 1981, Lievrouw and others, 1987)

- to trace the evolution of scholarly communities (e.g. Small, 1973; Garfield, Malin & Small, 1978; Small, 1988)

- to evaluate scholarly contributions (e.g. Latour & Woolgar, 1979; Garfield, 1985; Todovar & Glanzel, 1988)

- to study the diffusion of ideas (e.g. Winstanley, 1976; Paisley, 1984; Beniger, 1988)

This thesis is concerned with the first of those research questions: how to characterise the information policy scholarly community? The next section briefly reviews previous work which has tackled similar questions in other specialties

2.1.2 Previous bibliometric studies of specialties

A variety of approaches to bibliometric studies of scientific and other specialties is evident in the literature. Seven fairly typical examples are referred to in Table 2 1 on the next page to illustrate the range of approaches taken.

In his study of acupuncture, Haiqi uses a simple bibliometric approach to characterise the serials literature. The research aims are relatively modest: to provide a general picture of the current state of the researched literature on acupuncture. Haiqi notes that the results of the study "might give the working physician, researcher and librarian an idea of how the literature of acupuncture is distributed" (Haiqi, 1995:114). The paper examines the distribution of published articles by number of authors, language, geographical origin, type of therapy, and identifies the concentration of articles in specialist and

multidisciplinary journals. Very little attempt is made to contextualise the findings, either in relation to the needs of practitioners, information professionals or policy-makers. Sadly, examples of this type of study are ubiquitous, especially in library and information science journals, and have been dismissed as examples of "dry-as-dust empiricism" (Woolgar, 1991).

Table 2.1: Published bibliometric studies of academic specialties

Specialty	Research strategy	Reference
Acupuncture	Simple bibliometrics	Haiqi (1995)
Women's studies	Simple bibliometric analysis Content analysis	Cronin, Davenport & Martinson (1997)
Information science	Author cocitation analysis	White & Griffith (1981)
AIDS research	Bibliometric coupling	Small & Greenlee (1990)
Superconductivity	Longitudinal analysis of shifts in Bradford core journals Journal-journal cocitation analysis	Brooks (1990)
DNA polymerases	Patterns of coauthorship	Stokes & Hartley (1989)
Scientometrics	Analysis of co-authorship relations Analysis of citation networks Epistemic networks (title co-word analysis)	Wouters & Leydesdorff (1994)

Cronin and co-workers offer a much richer picture of the literature in the emerging field of womens' studies. Distributions of bibliometric elements (numbers of authors, for example) are compared with the patterns obtaining in other fields and some effort is made to contextualise the findings; the low incidence of author collaboration is interpreted, for example, as being indicative of the "early ghettoisation and the weak institutionalisation of the field within higher education" (Cronin, Davenport & Martinson, 1997:132). The simple bibliometric data is further enriched by means of content analysis.

Many bibliometric studies of specialties adopt a more structural approach, using various cocitation measures. Author cocitation analysis (ACA) is a popular

technique for studying the intellectual and social structure of specialties and many studies have been undertaken (see Section 8.2.4, p.146) since the appearance of the seminal article by White & Griffith (1981) which pioneered the technique. This method rests on the assumption that two authors are probably related to one another (intellectually and / or socially) if they are frequently cited together. Author cocitation studies offer highly simplified visual representations of the specialty using clustering and scaling techniques. The method is controversial, however, and tends to polarise opinion: Edge (1977) is perhaps the technique's most outspoken critic.

Other cocitation relationships (besides those linking authors) have been used to investigate the static and longitudinal structure of specialties. The study of AIDS research by Small & Greenlee (1990), for example, used bibliographic coupling data to illustrate changes in AIDS research over a six-year period. Brooks (1990) employed a similar technique but at a higher level of document aggregation. cocitations between journals to investigate the structure of superconductivity research. Unusually, Brooks also used a simple bibliometric technique, Bradford studies, to illustrate how the core journals in this specialty had shifted over time. In their study of the DNA polymerase research literature, Stokes & Hartley (1989) examined the patterns of co-authorship within the DNA polymerase specialty to investigate its social structure.

There appears to be a trend in more recent studies to draw upon several techniques in an attempt to develop a more reflexive and comprehensive understanding of the nature of specialties. The study of scientometrics by Wouters & Leydesdorff (1994), for example, draws on an analysis of networks of document cocitation and author co-authorship and supplements these findings with an analysis of title word co-occurrences ('epistemic network analysis').

It has already been noted that the development of documentary data sources, notably *ISI's* series of electronic citation indexes, has been a major stimulus to the wider acceptance of bibliometrics. To an extent, this could be viewed simply as highly opportunistic behaviour with the suspicion arising that studies are being carried out simply because access to large-scale bibliographic data is so

easy. Bibliometric techniques can certainly offer some valuable information about a discipline or field of studies, but there are naturally limitations, serious enough for some workers to reject the use of these techniques outright (see, for example, Edge, 1977; Woolgar, 1991; Buckland 1991). White (1990), however, offers a robust defence of the use of bibliometric data in studies of specialties, especially in relation to the use of author cocitation techniques.

2.2 The social basis of knowledge production

The question of precisely what constitutes an academic specialty or discipline is not at all straightforward, as even a cursory examination of the educational literature reveals (Toulmin, 1972; Hirst, 1974; Whitley, 1982). Two quite distinct but obviously related perspectives are evident in the literature: notions based around knowledge structures' and around 'knowledge communities'.

The idea that human knowledge, meaning and understanding may exist in quite distinct forms in different fields has fascinated philosophers since the time of Plato (see McGarry, 1991, for an excellent review of this huge topic). An understanding of the different forms and structures that knowledge may take in various disciplines is obviously an issue of great practical significance to educationalists engaged in curriculum design as well as to information scientists and librarians. The early classificationists (Dewey, Bliss, Ranganathan) thought that it was it was possible to identify and map reasonably permanent basic structures for given subject areas. This was based on rationalistic analysis and the assumption that it was possible to attain a high degree of consensus in the wider scientific community.

A key problem in classification, now widely recognised, lies precisely in the instability and lack of consensus that deny attempts to neatly scope and define different areas of knowledge. Hjørland attacks the rationalist approach vigorously, arguing that the work of Ranganathan and the others is:

> *"... too little concerned with real, organic disciplines, their development, differentiation, integration and mutual interaction, and too much concerned with artificial ideas that can be combined and separated like a puzzle in a mechanical fashion"* (Hjørland, 1992:198).

While acknowledging that a major limitation of this thesis is a reluctance to tackle the knotty epistemological difficulties associated with information policy (see Browne, 1997a, for an exposition of the problem), it is anyway doubtful how far such an approach might lead: information policy is widely recognised as an umbrella term embracing a diverse set of activities which are essentially practical and problem-oriented. To make matters even more intractable, Kajberg & Kristiansson (1996) argue that the field lacks a coherent theoretical framework, although they acknowledge that it might be pre-paradigmatic in a Kuhnian sense. Kuhn, reflecting on how scientific disciplines become established, suggests that:

> *"... it is sometimes just the reception of a paradigm that transforms a group previously interested merely in the study of nature into a profession or, at least, a discipline. In the sciences ... the formation of specialised journals, the foundation of specialists' societies, and the claim for a special place in the curriculum have usually been associated with a group's first reception of a single paradigm"* (Kuhn, 1962:19).

Interestingly, Kuhn notes that this need not necessarily be the case, citing medicine, technology and law, fields where the principal *raison d'être* is external social need, as possible exceptions to the rule. Does information policy sit here too?

An alternative perspective for understanding what constitutes a specialty or discipline is to view the field primarily as a social rather than an intellectual construct—as a knowledge community, a social network of like-minded people

bound together by common values and beliefs (Merton, *passim;* Toulmin, 1972). The two concepts are of course closely inter-related[7], as Hirst implies:

> *"...what then is one to make of the rather emotive term `discipline'? In that it suggests a tightly knit conceptual and propositional structure it would seem to apply most readily to a form of knowledge ... There is perhaps a tendency now to use the term in connection with an area of research and university teaching which professionals recognise as focusing on a large enough body of logically inter-related truths, theories and problems to justify its consideration in relative isolation from other matters"* (Hirst, 1974 97).

It would seem then that the attitudes, activities and cognitive styles of groups of researchers are closely bound up with the characteristics and structures of the knowledge d mains with which such groups are professionally concerned (Becher 1990). The social dimension can be very powerful, however. Geertz, an anthropologist, draws attention to the powerful cultural aspects of disciplines and the role of icons in establishing and projecting a carefully presented image—the chemist's desk with its three-dimensional models of complex molecular structures, the anthropologist's walls adorned with colourful tapestries, and the mathematician's chalkboard with its scribbled algebraic symbols (Geertz, 1983)[8]. These are not entirely facetious observations: Becher argues that "the tribes *[sic]* of academe define their own identities by employing a variety of devices geared to the exclusion of illegal immigrants" (Becher, 1990 24 The point here is that the notion of `specialty' is exceedingly soft and elastic, and certainly includes a major sociological dimension.

A recent article by Hjerland & Albrechtsen proposes a new information science paradigm—domain analysis—which states that "the best way to understand information in information science is to study knowledge-domains as thought or discourse communities" (Hjerland & Albrechtsen, 1995:400). Drawing on a wide range of material from educational research, psychology, linguistics, and the philosophy of science, the authors argue that much greater sensitivity is needed in information science both to the differences between knowledge structures and the specific social and cultural characteristics of user communities.

[8] On the basis of limited observation, I would assert that the information policy specialist may be similarly be distinguished by the ubiquitous framed political cartoon!

Bibliometrics cannot illuminate such a dimension, except very indirectly. One of the most important contributions of bibliometrics to the study of specialties has been the refinement of the concept of the `invisible college'. The term `invisible college' was originally coined in the seventeenth century by members of the Royal Society of London to emphasise the fact that they were geographically close together and shared common scientific interests, but lacked a building of their own (Lievrouw, 1988). The work of Price, notable *Little Science, Big Science*, brought the term back into common academic usage—but this time to signify informal networks of academics sharing similar interests, each with their own institutional base, but now separated by (often vast) geographical distances.

The concept of the invisible college speaks very clearly to the research questions set out in the previous Chapter. It emphasises scientific communication which, at least in part, is mediated through the medium of the refereed journal and suggests that bibliometric tools might be used to explore the cognitive and social links between members of the college. There is however an ambiguity which needs to be addressed first. Lievrouw notes that the invisible college is "typical of constructs that describe processes yet are founded on the study of structures; the ambiguity surrounding the use of the term is symptomatic of the confounding of structure and process in the study of scholarly communication" Lievrouw, 1990 59 . In other words, structural data may indicate the presence of communication relationships but they do not in themselves reveal the *nature* of those relationships.

2.3 Modes of knowledge production

The previous sections in this Chapter considered how bibliometric techniques might be used to develop insights into the intellectual and social structure of specialties. It was noted that the research questions addressed in this thesis relate essentially to the nature of scholarly communication. It is possible, however, to locate the findings of this thesis in an even broader context; as an example of `knowledge production' in a critical area of public policy. A recent book, *The New Production of Knowledge*, arising out of research at the Science Policy Research Unit (SPRU) at the University of Sussex (Gibbons and others,

1994), argues that knowledge production is shaped not only by scientific values but by wider social, economic and cultural imperatives. The authors perceive a major shift in contemporary societies in the way that specialised knowledge is created, and use the labels 'Mode 1' and 'Mode 2' to emphasise the transformation that is taking place. Some of the key differences between the two modes are summarised as Table 2.2:

Table 2.2: Modes of knowledge production

Mode 1	Mode 2
• disciplinary	• transdisciplinary
• highly institutionalised	• organisationally diverse
• knowledge generated within a disciplinary context	• knowledge produced in the context of application
• hierarchical	• heterarchical and transient
• peer review	• socially accountable
• codified knowledge	• tacit knowledge
• permanent	• transient
• fundamental and applied distinct	• flux

(adapted from Gibbons and others, 1994).

The issues raised in the book are profound. In many specialties, it is argued, research is no longer the exclusive preserve of the traditional university setting. Disciplinary and institutional boundaries are rapidly dissolving as broadly-based teams of researchers come together, often on an *ad hoc* basis, to tackle more complex social, technological or environmental problems. In other words, research agendas are increasingly being shaped from outside by political, economic and social stakeholders and knowledge produced within the context of application rather than for its own sake[9]. It follows that Mode 2 knowledge is generated across a wider range of institutional settings (universities, think tanks, private firms, professional associations) and that the results of Mode 2

[9] European Union research funding policy, especially in the context of the framework programmes, actively encourages research activity which transcends nationality, sector and discipline, although not necessarily in a way which all commentators regard as being economically or socially beneficial (Mahon, 1997).

research are far more highly contextualised than hitherto. The shift from Mode 1 to Mode 2 knowledge production has other, wide-ranging, implications. The traditional mechanisms for ensuring research quality, for instance, such as peer review, become less relevant as the emphasis shifts to social utility and accountability, and previously meaningful distinctions between 'pure' and 'applied' become less useful.

The work of Gibbons and others is highlighted here because it offers another framework for organising the findings of bibliometric studies at the level of specialties and disciplines. It suggests that particular attention should be paid to the issue of research collaboration (see Section 6.4.2, p.113) and also to the extent to which research activities in a given specialty are concentrated in just a few institutions or are more widely dispersed (see Section 6.4.3, p. 116). The nature of knowledge production in information policy is revisited in Chapter 10 where the empirical bibliometric evidence is discussed in the context of the ideas proposed by Gibbons and others.

Chapter 3: The field of information policy

> *"In a very real sense, information policy has come of age as a subject for study"*[10]

3.1 Introduction

This Chapter is a selective review, or meta analysis, of that small subset of the information policy literature which deals, introspectively, with the nature of its own subject. This is intended as a context for the work that follows: since, in the words of Parsons, "when we engage in meta analysis we are considering the methods and approaches used in the study of public policy and the discourse and language which it employs" (Parsons, 1995:1).

This brief review necessarily paints an incomplete and provisional picture: the literature surveyed is recent in origin (the earliest reference is Weinberg, 1963) a fact which in itself may suggest something of the immaturity of the field. Much of the material presented here is based on the author's editorial inputs to *Understanding Information Policy* (Rowlands, 1997), the published Proceedings of a British Library-funded workshop held at Cumberland Lodge, Windsor Great Park, in the summer of 1996.

3.1.1 The development of information policy

Information policy is just one of the many different types of policy decisions that governments and organisations make. In relation to other areas of public policy (e g. health, transport, education), information policy is, however, a very recent phenomenon. Burger (1993) traces the emergence of a systematic approach to information policy-making to the early 1960s and to the influential Weinberg Report on scientific and technical information transfer in the USA (Weinberg,

[10] Burger (1993:3).

1963)[11]. During the 1960s and early 1970s, the policy issues surrounding scientific and technical information and its use assumed a particularly sharp focus against the geopolitical backdrop of the Cold War and the race for the domination of space (Dunn, 1982; Hill, 1994). The subsequent emergence of large-scale data processing as a significant force in society and the growing volumes of information which were being held in machine-readable form by the early 1970s necessitated a policy response which went far beyond the boundaries of the scientific and technical community. For the first time, problems in areas as diverse as privacy, freedom of speech, secrecy, access to government information, and national security were being brought to the attention of policy-makers under a single integrating construct: the power of the computer to radically change society (Trauth, 1986). However, Burger argues that the literature of the time (the early 1980s) did not yet reflect "an identifiable field within the broad realm of policy analysis ... There had been discussion of these issues, to be sure, but no thread sewed these various policy fragments into a recognisable garment" (Burger, 1993:3).

The fact that information processing now offered a focal point for the discussion of such a wide range of content-related issues did however point out major gaps in conceptual thinking and public policy, and a vigorous debate about National Information Policies (NIPs) ensued which carries on to the present day (Oppenheim, 1996). Increasingly during the late 1970s and early 1980s, information was becoming recognised as a critical national resource, one which, if properly managed, offered the possibility of delivering economic, social and cultural benefits (Bushkin & Yurow, 1979; UNESCO, 1981; Cabinet Office, 1984). Possibly the most influential report, and the one which best captures the spirit of the times was the Nora-Minc Report. This report stressed the future importance of information technologies in French society and the economy and powerfully articulated the need for government to adopt a more positive role in managing change (Nora & Minc, 1980). Crucially, the report was commissioned by and reported to the President of France.

[11] Bennett (1987) and Rosenberg (1982) provide much useful historical material relating to the emergence of information policy as a focus of public policy concern in North America in the early 1960s.

More recent technological developments, notably in telecommunications and the widespread availability of personal computers, have simply accelerated processes that were already underway by the late 1970s. The influence of information technologies upon information policy (and vice versa) are profound. The case of data protection in the UK offers a good example of the extent to which information policy may be shaped by technological forces. Aldhouse (1997) traces the series of events that led to the UK's Data Protection Act in 1984. He shows how the legislation was originally framed by the need to regulate international flows of automatically processed information pertaining to individuals. As he points out, its motivation was primarily commercial and the issue a largely technical one, "it might as well [have] concern[ed] the thickness of plate for steam boilers". Aldhouse then charts how perceptions of that issue have shifted, particularly in response to the intensification of computer power and new applications which are having profound social consequences. He concludes that "the use of information technologies [has] become so pervasive an issue that information policy is [no longer] readily distinguishable from general social policy and notions of communitarianism and good citizenship".

The increasingly pervasive nature of information technology in modern societies, not least the emergence of a highly developed tradeable information sector, has led to a further broadening of the scope of information policy in the 1990s. Moore (1997) touches on various aspects of a grand design that politicians and journalists call the Information Society. In his analysis of the underlying issues behind the Information Society ideal, Moore identifies fear as a major component. fear borne out of poor economic performance, globalisation, structural employment, and North American cultural imperialism. He argues that Information Society policy is concerned with modernising the telecommunications infrastructure, promoting industrial and commercial competitiveness, re-skilling the workforce, promoting social cohesion, extending democracy by making governments more open and accountable, and contributing to cultural development. This is a breathtakingly ambitious list of policy objectives and it is rather surprising that so little public or political debate has fastened onto the subject, in the UK at least.

Information policy invites a range of opinion and Mahon (1997) offers a more sanguine viewpoint, characterising the Information Society as a convenient phrase for collecting together all sorts of issues and concerns in an apparently integrative framework so that politicians "can create a warm feeling in audiences". He suggests that in reality it has confused the debate, which ought really to be focused on practical issues of information availability and access. Webster, a British sociologist, is even more polemic:

> *"What I find most noteworthy is that Information Society theorists, having jettisoned meaning from their concept of information in order to produce quantitative measures of its growth, then conclude that such is its increased economic worth, the scale of its generation, or simply the amount of symbols swirling around, that society must encounter profoundly meaningful change. We have, in other words, the assessment of information in non-social terms - it just is - but we must adjust to its social consequences. This is a familiar situation to sociologists who often come across assertions that phenomena are aloof from society in their development (notably science and technology) but carry with them momentous social consequences. It is demonstrably inadequate as an analysis of social change"* (Webster, 1994:140).

While the problems which Moore identifies as driving Information Society policy are real and pressing, there is a real danger that politicians are guilty of over-determining their ideas, basing them upon unsupported assumptions about the effects of information on society. As Braman (1991) points out, these assumptions may be "so diffuse as to be inoperable ... The free flow of information, international co-operation, and peace are among the driving principles that have been offered to justify information policy".

3.1.2 The scope of information policy

Even this very brief review of developments in information policy must suggest that a succinct definition of the field is likely to be problematic. Parsons (1995) wisely observes that public policy fields tend to be defined by the problems and

issues that society faces at a particular point in time, not primarily by reference to some rigid epistemological framework. This view is obviously one which is shared by Hernon & Relyea, who define information policy as:

> *"... a field encompassing both information science and public policy [which] treats information as both a commodity—adheres to the economic theory of property rights—and as a resource to be collected, protected, shared, manipulated and managed. Although the literature often relates to information policy in the singular, there is no single all-encompassing policy. Rather information polices tend to address specific issues and, at times, to be fragmented, overlapping and contradictory"* (Hernon & Relyea, 1968:176).

Some writers, notably Burger, seem very reluctant to define the nature and scope of information policy at all, for fear of being labelled reductionist. However, since the intention of this thesis is to attempt to empirically map out the territory occupied by information policy scholarship, some working definitions of scope are needed. In order to provide a nomenclature to assist with search formulation and subsequently reveal the main disciplinary characteristics of the field, a subject classification scheme was essential. As might be expected, attempts to scope the issues and concerns which rightly belong under information policy abound in the literature. No *entirely suitable up-to-date* scheme could be found in the literature so a faceted classification was developed, freely adapting existing (unfaceted) schemes by Chartrand (1986), Milevski (1986), Rowlands & Vogel (1991) and Kajberg & Kristiansson (1996). The scheme used to classify information policy topics in this thesis is shown overleaf (as Content analysis frame 15).

Content analysis frame 15: Subject analysis of article

10	*General articles on information policy*
11	*Theoretical aspects of information policy*
12	*National and international information policies*
20	*Information infrastructure policies*
21	*Research & development (including STM policies)*
22	*Libraries, archives and public records*
23	*Telecommunications, broadcasting and superhighways*
24	*Information technology policies*
30	*Information management in government*
31	*Collection and acquisition of information resources*
32	*IRM in government: policies and practice*
33	*Information systems, clearinghouses and dissemination*
40	*Information access and control*
41	*Freedom of access to information*
42	*Confidentiality and personal privacy*
43	*Information control on grounds of national security*
50	*Information industry policies*
51	*Information standards and protocols*
52	*Copyright, intellectual property and information law*
53	*Regulation of the information industry and information markets*
54	*Trade in information services and transborder data flows*
55	*Public-private relationships in the information industry*

No attempt is made to offer any theoretical justification for the scheme: it serves its purpose simply as an *ad hoc* tool for exploring the range of concerns and disciplinary perspectives commonly addressed in information policy. It does, however, draw on the work of previous authors who have scoped the field and attempts to synthesise their thinking. Particular efforts were made to provide a scheme which reflects the recent emergence of interest in infrastructural policy areas (e.g. the construction of information 'superhighways') and in the structure and regulation of information markets. I would like to thank Dr Tamara Eisenschitz (City University) for her helpful comments on early drafts of this scheme.

Even a cursory examination of the scope of information policy, as normatively defined above, suggests that a number of disciplinary approaches and perspectives might be likely to be encountered in the literature. Indeed, information policy is frequently asserted to be highly interdisciplinary: Braman (1989), for example, claims that more than 40 academic fields deal with information policy, although she does not identify them. Burger (1993) argues

that the core disciplines which inform information policy are economics, law, political science, public administration, sociology, public policy, management science and information science.

The argument that information policy is highly interdisciplinary is a persuasive one, even though there is little empirical evidence to support the claim. The disciplinary nature of information policy raises some difficult questions. Where, for example, does information policy sit in relation to library and information science? Is the natural home for information policy within the LIS tradition, or does it share more features in common with broader public policy? Or are there unique characteristics of information policy which legitimate treating it as a field of studies in its own right?

3.2 Some unique characteristics of information policy

It would be quite possible, if premature, to enter here into a long debate about the relationship between information policy studies and the wider policy sciences. Browne has argued that the information policy literature is isolated from the mainstream of public policy research and that this is unhelpful (Browne, 1996). Others would certainly agree, perhaps going a stage further by noting that the information policy community has also missed the boat in relation to some important developments in the wider social sciences (Turner, 1997).

The question of the relation of information policy (with its roots in a highly positivist library and information tradition) to other, more established fields of public policy is an interesting one. In raising the question of what is *different* about information policy, I am in fact raising two sets of issues: not simply is it different in kind from, say, welfare policy, but, in thinking about information policy issues to what extent are we currently drawing on models from within information science as opposed to the policy sciences? What might be the contributions that these two very different traditions could bring to a unifying approach to information policy studies?

It is useful to turn at this point to the writings of Sandra Braman, one of the most consistently rewarding authors on information policy. In a 1990 paper, Braman identifies the factors that for her differentiate information policy from other policy issues. Information policy, she argues, is 'different' because:

- it is a relatively new area of policy concern
- it involves unusually diverse groups of players
- decisions about information can have an enormous impact on events and policies in other areas - the reverse is true to a much lesser extent
- information does not fit into the traditional categories employed by policy analysts
- information policies made at very different levels of the political and social structure, from the local to the global, are remarkably interdependent

These differences are all, of course, relative. They may not be unique (i.e. exclusive) characteristics in the strictest sense, but they certainly seem to be atypical of other areas of public and organisational policy, where the issues tend to be relatively more clear cut and the affected parties less diversified.

Braman pursues the notion that information policy is unusually complex in a public policy sense and generalises her argument to suggest that four general characteristics of information policy, regardless of setting, are major sources of that complexity. These sources are conceptual, informational, structural and orientational in nature.

3.2.1 Conceptual problems in information policy

Conceptual problems arise from the fact that the rate of technological change is so fast that it is outpacing our ability to keep up (see Pye, 1997, for a graphic picture of the problems of runaway technology); that regulatory structures and

values sometimes conflict (Eisenschitz, 1997; Worlock, 1997); that policy goals and regulatory goalposts are not static and may shift considerably (see Aldhouse, 1997, on the evolution of data protection policy). Further conceptual problems and dilemmas abound in the information policy literature, even over the interpretation of basic terms like 'policy' and 'strategy' (Braman, 1989; Allen & Wilson, 1997). Ominously, there appears to be neither consensus nor even much in the way of recent debate about the meaning of the term 'public interest'.

3.2.2 Informational problems in information policy

Braman argues that two types of problems emerge from the roles that information *itself* plays in the policy-making process: those that stem from a lack of information, and those that stem from the effects of skewed information. A review of the information-seeking behaviour of policy-makers by Rowlands & Strachan (1997) suggests that policy is quite often made in spite of the fact that the knowledge needed to fully support a decision is unavailable. Braman (1990), quoting many examples from US policy-making, finds that policy decisions are often based on "bad, outdated, non-academic, unscientific studies" or are made in an "informational vacuum". More excusably, perhaps, she finds also that decision-makers repeatedly claim themselves to be "incapable of understanding technological development well enough to regulate it".

The other source of confusion arises from what Braman calls the 'informational deficit'. This argument says that the information available to policy-makers tends to be skewed in favour of what is politically and ideologically acceptable: she makes particular mileage out of the conceptual slippage between the image of the marketplace and its reality.

3.2.3 Structural problems in information policy

Structural problems are those that arise out of conflicts or contradictions between different policies or between different elements of the policy-making machinery: a common theme in writings on national information policy. It may be resolved into two components. The first is that seamless co-ordination

between, for example, government departments, may not be possible in information policy terms because the goals and world vision of the parties involved are distinct and not necessarily capable of being resolved. A second, related, point is that the sheer complexity of information policy results in a reduction of the capability of the state to make policy coherently and effectively. The fact that so much public information policy is latent rather than explicit (see, for example, Martyn, 1990) becomes a real structural problem: compared even with stated policy, unstated policy is capable of many differing shades of interpretation and realisation. Hill cites the old joke, still to be heard in Britain, that "Our information policy is not to have a policy" (Hill, 1990:3).

3.2.4 Orientational problems in information policy

Finally, Braman addresses the big question of what information policies are designed to achieve. Without a clear sense of orientation, such fundamental questions as: What kind of social organisation do we want? What kind of informational organisation best serves that kind of society? What kind of regulatory structures best build and sustain that kind of informational organisation? all become superfluous. These are clearly highly political issues, with a capital P.

The uniqueness of information policy seems to be a function of its complexity. The encouraging aspect of Braman's paper is the (implicit) suggestion that, at least in part, that 'complexity' is simply a result of our confused thinking (Rowlands & Turner, 1997).

3.3 The normative structure of information policy

Maybe one key to unlocking that complexity is to first acknowledge our own positions as policy analysts and policy actors: no-one involved in information policy can really believe themselves to be totally objective and free of assumptions and prejudices. Progress in information policy studies is critically dependent on finding ways to make those assumptions and prejudices more transparent and challengeable. Information policy, like all aspects of public

policy, is deeply embedded in a political and cultural context from which it derives its values and its sense of direction. Some would even argue that information policy is fundamentally about the exercise of power by the state (Braman, 1991). In this context, information policy is just one strategy of many by which the state establishes and consolidates its authority.

The role of norms and values in information policy-making is a rather neglected area, although Overman & Cahill (1990) offer a very insightful programmatic article on this topic. They argue that the normative structure of information policy urgently needs to be explicated both in research and practice. This is an important point and is illustrated here by means of an example. Classically, the genesis of a public policy first involves the recognition of an issue. Consider, for example, the following sequence:

ISSUE	People sleeping on the streets
PROBLEM	Homelessness
POLICY RESPONSE	New forms of policing

What counts as a problem and how a problem is defined depends very much on the way that the policy-makers perceive it. A problem has to be defined, structured, located within certain boundaries and given a name. The mechanics of this process prove crucial for the way in which a policy is addressed to a given problem. If we see people sleeping on the streets as a problem of vagrancy, then the policy response may be framed in terms of law enforcement and policing. We might also view the same issue as an indicator of social deprivation or a sign of failure in other policy areas such as community care; in which case the policy response will obviously be very different (we might provide low cost housing, for example, or appropriate mental health care).

Information policy issues are often equally ambiguous. Consider, for example, the recent trend towards the electronic delivery of public services. Local government in London has been quick to embrace electronic communications, using e-mail, the Internet and the World Wide Web. Thirteen London boroughs now have e-mail, and more than twenty boroughs have or are planning their own

official World Wide Web sites. These Borough Web sites contain a wide range of information, including 'What's On' guides of borough activities, council reports and plans, tourist information, and links to local community and voluntary groups. Many boroughs use innovative forms of electronic communication to provide information about local services: examples include public libraries with computers hooked up to the Internet; electronic kiosks displaying pages of local information in public places like supermarkets; and videotelephony for hearing impaired people.

Simultaneously, it is possible to see these developments as informing and empowering local communities; extending democracy; as instruments to reduce administration costs and head count by re-engineering local government services; providing a stimulus to the information industry; a new weapon in the councils' public relations armoury; or merely as gesture politics, pandering to the IT fetishists. The truth probably lies somewhere (everywhere?) on this list, but it is not immediately evident, especially as the waters become clouded by practical, legal and ethical issues arising in the implementation of these services.

Economic arguments over whether information should be regarded as a tradeable commodity or a public good abound in the library and information science literature. Sometimes, if rarely, both characteristics may be accommodated by means of differential pricing structures (see, for example, East, 1977). The debate on this topic is extensive, but personal views tend to be fairly fixed and unmoveable. Eisenschitz (1997) points out that the creation and development of information goods and services requires that the investors be rewarded for their efforts. This means conferring property rights and enabling a marketplace to develop. On the down side, these rights may potentially erect a barrier to the world of ideas for those who are unable to pay. This in turn necessitates a further set of public policies, such as universal service, investment in libraries, research & development, and education which deliberately distort the information marketplace in favour of the wider public interest. Questions of this kind are ubiquitous in the professional library and information literature.

Aside from monetary considerations, access to information and knowledge is also a function of power structures. In fact, there is an inherent tension in the notion of access quite distinct from the economic arguments outlined above. In many situations, the widest possible access to information is seen as a 'good thing': health promotion, information about school performance, consumer information, access to local government records, being classic examples. In other circumstances there are real problems in allowing unrestricted information access: policy advice to ministers, personal information, information which might be prejudicial to national security or a firm's commercial position. Restrictions on the free flow of information are sometimes essential, but there is always a danger that powerful forces in society will constrict these flows for their own advantage.

Thus, information policy is inherently value-laden, especially so at national and international levels, and policy-makers have daily to face a very difficult balancing act, negotiating an acceptable compromise between conflicting objectives, values and interests. These are often diametrically opposed, as the following examples show:

- market-led *versus* state-led visions of the Information Society (Moore, 1997)
- freedom of expression *versus* rights of personal privacy (Aldhouse, 1997)
- the monopoly functions of patents and copyrights *versus* their informational aspects (Eisenschitz, 1997)
- the philosophy of open government *versus* retention of crown copyright and restrictive licensing practices (Worlock, 1997)

Overman & Cahill (1990) offer a useful framework for understanding the normative basis of information policy, as exemplified in US Federal legislative, executive and judicial decision-making over several decades:

Table 3.1: Overman & Cahill's information policy values

Access and freedom	If democracy is to work, then people need to be well informed. Although rights to access information are enshrined in legislation in many countries there are significant tensions around the questions of access, security, and freedom of information.
Privacy	This is a widely accepted value in Western-style democratic societies although there are tensions which arise out of the needs of government and society on the one hand, and on the other, the individual's rights to privacy in their affairs.
Openness	This is about open government and the right of people to know about the processes of decision-making in government.
Usefulness	With the rise of the administrative state, governments began to collect records. 'Usefulness' refers to the idea that the worth of information is determined with reference to the use to which it is put. A key issue is who decides what is useful and, therefore, what is to be collected and stored.
Cost and benefit	This assumes that information has economic value, costs and benefits and raises problems of reconciling commercial interests and the public interest in the matter of the information collected by government.
Secrecy and security	Secrecy and security are two high profile values with significant issues around the question of the amount of latitude to be given to government officials.
Ownership	Intellectual property addresses not the ownership of tangibles but the form and expression of ideas through patents or authorship of texts of any kind. Here, issues of commercial interest can collide with the needs of individual users and society.

This value set provides a very useful tool for understanding what underpins information policy both as an area of scholarship and an area of practice. Seen from this perspective, information policy might be seen as a strategy for embodying and acting upon these fundamental values.

Values provide rather a slippery context for information policy, but they raise an important question: what is the nature of the overall framework within which information policy-making takes place? As Kristiansson (1996) points out, information policy-making activities have historically tended to focus *ad hoc* on specific issues and problem areas such as research & development, information market development, freedom of access to official information, legal aspects such as privacy, copyright and intellectual property rights. The legacy of this approach has been a fragmentation of policy-making responsibilities, institutions and discussion fora (see, for example, Oppenheim, 1996). The evident lack of co-

ordination of these disparate but obviously related policy strands is something of a pre-occupation among writers on national information policy. It may well be argued that information policy is so loose and broad a concept that it cannot and perhaps should not be accommodated within a single ministerial portfolio. What cannot be escaped, however, is that there appears to be no explicit overarching framework within which issues and conflicts may be resolved.

3.4 Towards an information policy regime?

Eisenschitz (1997) argues that it is important that we work out commonly understood criteria for assessing policy in the information field: we currently neither have common measures nor a common baselines of assumption. If the Information Society concept is to become a reality, she argues, a more unified approach to information handling and a more consistent regulatory framework will be required. Worlock (1997) elaborates much the same point in his analysis of the ineffectiveness of UK tradeable information policy. He notes that there are real practical issues around defining precisely what is a source or primary document, and even more difficult, in defining precisely what constitutes 'value added'.

Consensual approaches to thinking about a particular policy issue area in ways that provide a basis for decision-making are called 'regimes'. A regime may be thought of as a normative framework that is less rigid and formal than a legal system but nonetheless serves to provide a common understanding binding all parties together. Since information policy is relatively new area of public policy concern, an information policy regime may only just be evolving. While working policy-makers cannot put aside immediate problems until the larger theoretical issues have been resolved, it is possible to approach problems with an awareness of their theoretical context and large-scale socio-economic, political and cultural impact. The final thought goes to Sandra Braman:

> *to be information-literate in the twenty-first century will be to understand that information technologies, messages, institutions, and effects within*

the context of ongoing explorations in both epistemology and the sociology of knowledge (Braman, 1990:77).

Chapter 4: Information policy research methods[11]

> ***"The London underground map is an example of constructing a map which enables us to use this highly complex system. However, it is more than that. For many, including the author, the map has come to form a mental image of London. The danger for tourist and resident alike is that Beck's map becomes our image of London"[12]***

4.1 Introduction

This Chapter identifies and describes the main research tools, methodologies and perspectives that have been used to create new knowledge in the field of information policy. The materials consulted for this review are drawn mainly from the library and information science literature but, for reasons which will become clear, it has also been necessary to look beyond the boundaries of the library and information science literature in a search for *potentially* applicable models and frameworks.

This Chapter is concerned primarily with the study of information policy from an academic perspective, motivated chiefly by curiosity and the desire for greater understanding. It does not claim to provide specific insights into the 'black box' of policy formulation within the executive or legislative branches of government. It certainly does not conclude with recommendations for what information policy *should be*. Its intention is directed instead towards a better understanding of what information policy *is* and what tools, frameworks and methodologies are currently available for its systematic and critical analysis.

A more specific objective of this literature review is to build a platform for the content analysis of a bibliography of information policy articles (see Chapter 9); this is achieved by developing a series of coding frames that are later used to

---

[11] This Chapter was originally published as 'Understanding information policy: concepts, frameworks and research tools', *Journal of Information Science 22(1)* 1996, 13-25.
[12] Parsons (1995:61).

map the distribution of various research strategies and methodologies across the serials literature. As in the previous Chapter, these coding frames are presented and discussed in the body of the text and consolidated as Appendix D.

4.1.1 The basic toolkit

The first problem in considering the nature of the information policy serials literature is to define what actually constitutes a 'research paper'. Peritz, in a bibliometric study of library science research, used the following working definition:

> *"Research is an inquiry which is carried out, at least to some degree, by a systematic method with the purpose of eliciting new facts, concepts or ideas"* (Peritz, 1980).

This definition carries no implication of quality, relevance or generality; so long as its procedures are systematic and objective, and its aim is to reveal something hitherto unknown, an article qualifies here as a 'research paper'.

Content analysis frame 16: Type of article

1	*Based on original research*
2	*Not research-based*

In applying these categories for the purposes of content analysis, Peritz' broad definition of a 'research paper' needed to be translated into specific operational terms for each method (see Table 4.1 overleaf):

Peritz identified six main types of research method in library science: Theoretical; Surveys and Experiments; Secondary Analysis; Historical Research; Design of New Methods and Procedures; and Descriptive Bibliography. Apart from these main categories, Peritz noted that a bewildering array of other methods were in evidence, ranging from chemical analysis to jurisprudence.

Table 4.1: Operational criteria for characterising original research

Research method	Operational criteria
Historical research	• use of primary data sources
Comparative studies	• evidence of critical analysis and comparison of documentary or other sources leading to significant new insights
Survey methods	• use of surveys, experiments, bibliometric studies, content analysis or other quantitative tools
Case studies	• evidence of critical analysis and comparison of case study findings leading to significant new insights
Secondary data analysis	• data re-analysed logically or statistically • data from different sources integrated in a way that reveals new insights
Literature review	• evidence of critical analysis and comparison of documentary sources leading to significant new insights
Development of new theoretical frameworks	• considered as a 'research paper' even when the article contains no evidence of systematic method, if framework is (a) sufficiently novel, and (b) generally applicable

In developing a coding frame for the analysis of information policy articles, a faceted classification scheme has been adopted, closely following the work of Jarvelin & Vakkari (1990, 1993). This separates research methods into two groups: *empirical* and *conceptual* strategies:

Content analysis frame 17: Research method

10	***Empirical research strategies***
11	***Historical research***
12	***Comparative studies***
13	***Survey methods***
14	***Case studies***
15	***Secondary data analysis***
20	***Conceptual research strategies***
21	***Literature review***
22	***Development of new theoretical frameworks***

The adoption of this content frame is intended to facilitate comparison with Jarvelin & Vakkari's extensive content analyses of core LIS journals over the period 1965-1995.

4.1.2 The aims of information policy research

Complexity runs as a leitmotiv through the information policy literature. Information supply, transfer and use take place within an environment which is in a constant state of flux, shaped by the unpredictable interaction of commercial, economic, technological, social and demographic forces (Hawkins and others, 1992). One of the most difficult and fascinating problems in information policy research is that public policy has an influence, directly or indirectly, on each of these sets of forces; hence, even such broad policies as those relating to education, open government or the funding of civil science may have dramatic implications for information availability and use. The complex and rather abstract nature of information policy means, however, that public understanding and consensus on such key issues as trading in government information or data privacy is less than it might be. The popularisation and wider dissemination of information policy research may have a useful role to play here in raising awareness of the issues and contributing to a more open public debate.

Despite, or perhaps because of the observed complexity, many analysts have argued for more `rational' approaches to the study of public policy, pointing, for example, to the need for more sophisticated indicators of economic and social impact, better forecasts, clearly expressed hierarchies of objectives and improved definition. As will be seen in this Chapter, these are some of the areas where the currently available tools for information policy analysis appear to be less than wholly adequate.

The notion that government can, by making policies, `solve' problems lies at the root of what has become known as the `policy sciences'. Back in the 1930s, John Maynard Keynes argued that if government was to have any chance of dealing with the problems of the day, it had to recognise the need for a more informed, theoretically-driven approach to governance. In the future, he predicted, it would be the ideas of economists rather than political interests that would shape decision-making (Keynes, 1936).

A more unified approach to the study of public problems and policy was initiated with the early work of Lasswell (1930, 1948); culminating in the publication of *The Policy Sciences* with Lerner (Lerner & Laswell, 1951). Lasswell describes the role of the contemporary policy scientist as:

> *"... an integrator of knowledge and action, hence as a specialist in eliciting and giving effect to all the rationality of which individuals and groups are capable at any given time. He is a mediator between those who specialize in specific areas of knowledge and those who make the commitments in public and private life ... Both the intellectual community and the community at large are beginning to acknowledge the indispensable place of the integrator, mediator, go-between"* (Lasswell, 1970:13-14).

It follows from these remarks that the policy sciences are contextual, multi-method and problem-oriented and that a wide range of research approaches might be expected to be encountered in the literature.

A fundamental distinction in the public policy literature is that between *policy studies* and *policy analysis*. These two types of research are differentiated by their objectives and by their relation to the policy-making process. Policy studies are studies *of* policy; they are motivated by curiosity rather than an explicit aim to shape the course of events. Policy analysis, in contrast, is research which actively seeks to influence the policy agenda. It is research *for* policy, and it usually involves the production of highly value-added information for policy-makers.

Gordon, Lewis & Young (1977) argue that the terms policy studies and policy analysis subsume five distinct types of research strategy:

Policy determination is concerned with how policy is made, why, when and for whom.

Policy content may involve either a description or a critique of a particular policy and how it relates to other, possibly earlier, policies.

Policy monitoring and evaluation examines the impact of policies and how they have performed in relation to their original objectives.

Information for policy is a form of analysis which is intended to feed directly into policy-making processes. This may take the form of detailed information and research or advice-giving.

Policy advocacy involves research and arguments which are intended to influence the policy agenda.

Content analysis frame 18: Analytical focus

Policy studies	
1	*Analysis of policy determination*
2	*Analysis of policy content*
3	*Policy monitoring and evaluation*
Policy analysis	
4	*Information for policy*
5	*Policy advocacy*

A recent and much-debated theme in the policy literature focuses on the interface between the 'two communities' of policy researchers and policy-makers[13]. The central issue here is what kind of rational analysis is compatible with the real world of decision-making in which there are conflicts over facts and values, means and ends, and in which there is considerable uncertainty? Quade (1976) concedes that policy analysis can never be an exact science[14] since its primary concern is to help a decision-maker "make a better choice than would

[13] Articles by Bearman (1988) and by Hammond and others (1983) suggest that information science methods can be applied to advantage to the study of how policy-makers use and filter information inputs during the early stages of the policy process (problem identification, agenda setting and policy formulation). Their findings suggest that there are fundamental obstacles to the effective use of scientific information in public policy-making and that 'rationality' in this context may be subservient to other factors such bureaucratic goals and political expediency.

[14] The belief in the possibility that decision-making in the public sphere can be made more rational is regarded as highly questionable by 'incrementalists' such as Lindblom (Lindblom, 1959, 1979). Even Lindblom acknowledges the need for some kind of rational inputs into decision-making, however.

otherwise have been made", thus aiming to facilitate an "effective manipulation of the real world".

The strategy which frames rational policy analysis is often expressed in terms of a cycle involving five key stages (formulation, search, forecasting, modelling and evaluation), a framework which maintains that a decision is the result of a series of logical steps. Quade (1976) associates each of these stages in turn with a different primary knowledge requirement:

Content analysis frame 19: Main objective of article

1	*Clarifying the problem*
2	*Identifying and screening the policy alternatives*
3	*Predicting the future environment and operational context*
4	*Modelling policy impacts*
5	*Ranking policy alternatives*

In Quade's scheme, the role of analysis is to facilitate a rational choice of means and ends, within the limits of what is possible given constraints in the `real world'.

Another way of thinking about policy-making is to locate it in a cultural or ideological setting. Richardson and others (1982) developed the idea that policy researchers, politicians and decision-makers each exhibit a variety of patterns, or style, in the way that they think about problems. One of the dimensions of this style is whether they have a tendency to anticipate problems or to react to events and circumstances as they arise.

Content analysis frame 20: Analytical style

1	*Reactive*
2	*Anticipatory*

4.2 Some dimensions of the information policy problem

Before looking in more detail at some of the concepts, frameworks and research tools which have been applied to the study of national and international information policy, some preliminary remarks are called for in relation to the object of study and its highly problematic nature. First of all, what is information policy? There are at least as many definitions of information policy as there are writers on the subject; in this section a few broadly representative examples will be used to illustrate some important points about the characteristics of information policy. Weingarten (1989), for example, defines information policy as comprising "the set of all public laws, regulations, and policies that encourage, discourage, or regulate the creation, use, storage, and communication of information". This is a broad and inclusive definition; it suggests that the fundamental role of policy is to provide the legal and institutional frameworks within which formal information exchange can take place. It implies that information policy addresses both political and bureaucratic goals (an interesting point, as these may not necessarily be congruent) and since public information policy emerges out of the machinery of government, its formulation, implementation and evaluation may be expected to be accompanied by the production of various documents: discussion papers, legislative drafts and memoranda. To a greater or (usually) lesser extent these may be available for purposes of study; an idea which will be returned to later in the section on tools and methodologies.

Another valuable insight into information policy is offered by Hernon & Relyea (1968), who underline its complex nature by noting that "although the literature often refers to information policy in the singular, there is no all-encompassing policy - rather information *policies* tend to address specific issues and, at times, to be fragmented, overlapping and contradictory". Certainly, the emergence of information processing as a significant force in society has been accompanied by an extraordinarily wide and diverse range of public policies; from the regulation of value-added and data services to the licensing of government data sets for commercial exploitation and legislation to protect personal data privacy. The extraordinarily complex and diverse nature of information policy is a common

theme in the literature (Malley, 1988; Gray, 1988; Suprenant, 1987; Moore, 1992) and one which poses serious difficulties for research design. Faced with what sometimes appears to be at best a 'fuzzy set' or at worst a set of unconnected or contradictory laws, regulations and policy statements, issues of scope and definition in the field of information policy enquiry become highly problematic. The situation may be messy, but it is inevitable given that the practical realities of information policy are much to do with the art of compromise: there are neither 'good' nor 'bad' policies but maybe there can be effective compromises between competing interests. Galvin (1994) states this idea with admirable clarity, noting that information policy-making comprises: "a fundamental enduring conflict among or between objectives, goals, customs, plans, activities or stakeholders which is not likely to be resolved completely in favor of any polar position in that conflict". This suggests that another feature of information policy is that it is, or at least that it should be, flexible, dynamic and responsive to changing circumstances. It also implies that the impact and outcomes of information policy decisions difficult to predict with any certainty, and that a key objective for information policy research might therefore lie in trying to set some kind of bounds and limits to that *uncertainty*.

Given the complexities and uncertainties associated with information policy, how widely should the scope of academic studies be drawn? Kristiansson (1996) argues that questions of information policy should be addressed at an appropriate spatial level, so as not to confuse macro- and micro-level factors.

Content analysis frame 21: Information policy scale

1	*Global*
2	*International*
3	*National*
4	*Regional*
5	*Industrial / sectoral*

A wide range of forces other than public policy are relevant to a consideration of the issues affecting the supply and use of information goods and services in society. These include, *inter alia*, the commercial strategies adopted by publishers and database providers, the behaviour of consumers, the influence of

pressure groups, and the structure and patterns of ownership within the information industry. One of the characteristics of public information policy is that it both *shapes* and *responds* to events and so can be regarded either as an independent or a dependent variable in scientific policy studies. As an *independent* variable, information policy can be analysed in terms of its impact and outcomes, both on the wider environment and on the political process itself. When it is viewed as a *dependent* variable, our attention is drawn to the environmental, cultural, economic and other factors which shape and guide policy and its implementation. This insight leads Burger (1993) to define information policy even more broadly and inclusively as those "societal mechanisms used to control information, and the societal effects of applying those mechanisms". Clear differentiation between the treatment of information policy as an independent or a dependent variable in analysis should be an important consideration in research design, but my observation is that little is made of the distinction in the serials literature.

There is an understandable tendency for writers on information policy to define the scope of their work narrowly and in a self-referential way; to discuss issues surrounding intellectual property rights or public library policy solely in terms of their implications for information professionals. In reality, however, it is impossible to detach information policy from its wider social and political context, a point which the Secretary of State's speech to the Library Association Umbrella Conference on 9 July 1993 underlines:

> *"... there are many major challenges facing the library world at the moment. To take a few of the more prominent ones: local government reorganisation; new further education funding arrangements; competitive tendering for public services ... all these issues stem from government policies originating outside the Department of National Heritage's field"* (Department of National Heritage, 1993).

These remarks suggest that there is, in effect, an *information policy hierarchy* comprising three levels: infrastructural, horizontal and vertical policies. Examples of infrastructural policies might include tax or employment law,

freedom of establishment, and education policy, which apply across society and which affect the information sector both directly and indirectly, providing a social and economic context for its activities. In contrast, horizontal information policies have specific application and impact across the information sector, such as the statutory provision of public library services, zero-rated VAT on books, or data protection law. Vertical information policies have specific application to a particular information sector, such as the geographic information community.

Content analysis frame 22: Information policy hierarchy

1	*Infrastructural policies*
2	*Horizontal information policies*
3	*Vertical information policies*

This hierarchical model is useful in that it locates information policies in the broader context of public affairs. This may be productive in itself, but it further extends the scope problem: it becomes very difficult to conceptualise information policy in its most inclusive and comprehensive sense, since this would need to include, as well as an analysis of narrow sector-specific policies such as intellectual property rights and information access policies, the wider contexts of educational, social and industrial policy.

However, all three levels must be of interest to the information policy analyst because of the inevitable degree of interaction between them. Indeed, the overlap between the three levels can sometimes seriously obscure our understanding of what 'information policy' means and what it actually consists of. Martyn makes the point that the absence of horizontal policies (in the above sense) does not necessarily mean that there is 'no policy'. Rather, information policy may be subsumed by infrastructural policy:

> *"... in effect, there is no real policy for the information industry, electronic or otherwise, other than that which conforms to the general Government policy towards industry, which is to allow free market forces to operate within a lightly regulated environment, the regulation being put in place when the need becomes evident and pressing"* (Martyn, 1992: 270).

An implication of the hierarchical model of public policy is that power, influence and decision-making in relation to information policy are inevitably scattered across different parts of government. A major conceptual problem therefore lies in defining an overall set of policy values and a framework for specific information policy actions. In other words, there is a need for the establishment of a *information policy regime* which is comprehensive, sensitive to new technology and responsive to the implications of the 'information age'. The lack of such a coherent framework is a relatively new theme in the information policy literature (Braman, 1990; Gray, 1993).

Moore (1993) has provided a useful and practical way of beginning to deal with this problem in the form of a two-dimensional space which focuses attention on the differing needs of industry, individual organisations and society (Table 4.2).

Table 4.2: The information policy matrix

	Industrial	Organisational	Social
Information technology	o	o	o
Information markets	o	o	o
Information engineering	o	o	o
Human resources	o	o	o
Legislation & regulation	o	o	o

Moore recognises and rejects as futile the often-rehearsed argument which calls for a single unified expression of national information policy. Instead, he argues that the effects of technological and other agents of change are so powerful, and the underlying issues so complex, that a more flexible approach to policy-making is needed. As well as offering a framework for the analysis of information policy objectives, the model makes it possible to identify gaps and areas where the formulation of policy objectives might be appropriate.

4.3 Theoretical approaches to the study of public policy

Political and social scientists have developed a number of models, frameworks and theories for analysing public policy, although few of these have been explicitly applied and tested in the context of information policy studies. As well as lacking theory, we lack a sound basis of empirical data relating to information

policies and their outcomes: indicator data relating to their impacts are either very difficult to obtain or are simply not available. Yet without a body of theory it is difficult to see how a more rigorous (possibly more quantitative?) approach to information policy analysis can be forthcoming: it has often been observed that there is nothing more practical than a good theory! Burger (1993) notes that information policy is: "complex because we have not yet developed ways of understanding it that are widely accepted and broad enough to encompass its range: we now rely on existing disciplines to inform us about the soundness of information policies". This may be an honest appraisal of our current difficulties in defining the scope and boundaries of the subject, but it offers little practical comfort.

Broadly speaking there are three sets of motivation for studying information policy problems; for scientific, professional, or political reasons. The scientific rational is motivated primarily to gain a greater understanding of how they originate, how they are developed and implemented, and what their consequences are for particular interest groups or for society as a whole. The scientific motivation seeks to understand policy, not to suggest what that policy ought to be. Clearly, however, information policy studies undertaken for professional or political ends have a different emphasis: here analysis has an applied orientation and is concerned with determining the most efficient or `best' alternative; the one that will yield the largest net benefit for dealing with a particular problem. This approach is concerned with achieving the `right' goal, with what policy *ought* to be, and cannot therefore be arrived at without reference to an ideological ('the market should decide') or normative ('the preservation of cultural heritage is paramount') position. As such, it is a value-oriented approach and one which should be located outside the scientific policy studies tradition, one which asserts itself to be rigorous, objective and value-neutral.

A useful working definition of the scientific approach to policy studies is offered by Dye (1984), who suggests that it comprises: "a description of the content of public policy; an assessment of the impact of environmental forces on the content of public policy; an analysis of the effect of various institutional arrangements

and political processes on public policy; an enquiry into the consequences of various public policies for the political system; and an evaluation of the impact of public policies on society, both in terms of expected and unexpected consequences". Within this tradition, Anderson (1990) identifies four dominant theoretical approaches: political systems theory (Easton, 1965); group theory (Latham, 1965); elite theory (Dye & Ziegler, 1990); and rational-choice theory (Downs, 1957). The applicability and potential value of these theories in the specific context of information policy is unclear (and beyond the scope of this thesis) but they possibly represent a fruitful line for future enquiry.

One of the features of information policy research is the number and diversity of academic disciplines which have, with some justification, laid claim to being the home discipline for studies in the area: such is the fragmentation of the field that Braman (1989), a researcher from a communications studies tradition, identifies more than 40 academic fields that deal with 'information policy'. Braman does not offer a list, but Burger (1993) suggests that the core information policy disciplines comprise economics, law, political science, public administration, sociology, public policy, management science and information science.

In trying to understand the highly distributed nature of information policy research, it is worth considering three factors: the multiple interpretations of the term 'information' by policy-makers in different economic, social and cultural contexts; the way in which information policies have developed over time as pragmatic solutions to specific problems; and the highly institutionalised nature of policy studies.

Traditionally, information policies have evolved in direct response to the emergence of specific technologies, such as print, telephony, radio or value-added and data services. Not surprisingly, the analysis of these policies has tended to fall within the domain of whichever professional information community was most directly concerned with the particular technology involved (as librarians, computer scientists, broadcasters or information scientists). In other words, while information policy has been *technology-driven*, policy research has typically been *discipline-bounded* (Trauth, 1986). The fragmentation of

information policy research is mirrored by a fragmentation of policy-making institutions. In Britain, for example, the Department of Trade and Industry is the lead agency for developing policy in relation to tradeable information, standards and intellectual property; while data protection is under the jurisdiction of the Home Office; legal deposit under the Department of National Heritage; and public records under the Lord Chancellor's Department.

If information policy studies are discipline-bounded, they are also to some extent institutionally-bounded; just like the policies themselves, research emerges from bureaucratic institutional settings which have their own distinctive goals, perspectives and agendas. Sometimes, academic studies are carried out to directly support the efforts of those engaged in policy formulation, DG-XIII of the European Commission, for instance. Such research is likely to be highly prescriptive and goal-oriented, designed to meet the immediate needs of policy-makers. Other studies are undertaken to meet the needs of the information industry, perhaps in the form of market studies and assessments, or (occasionally!) funded by a research council where longer-term, more fundamental objectives are sought. The linkages between research and research funding are important because they reinforce the process of fragmentation; information policy develops along different research fronts, informed by very different theoretical approaches; factors which have not been conducive to establishing consensus and agreement on central information policy issues and values.

While most observers would agree that there is currently little consensus over where the boundary lines of the information policy discipline should be drawn, there are even those who actively resist such an approach. Braman, for example, argues that `theoretical pluralism' is the most appropriate way to deal with phenomena and processes which unfold in different ways at different levels of the social and economic structure. Her argument offers two challenges to those engaged in information policy research: firstly, the need for a high degree of self-awareness of the assumptions and paradigms being brought to bear from within one's own disciplinary tradition, be that library and information science, telecommunications policy or media studies; and secondly, a reminder to beware

of the sweeping generalisations which inform much writing (and dare one say public policy) in this area. We should not simply accept that broad statements of policy intent, such as 'the protection of national security' are in themselves a sufficient justification for information control, particularly since these are too often only a guise for economic protectionism or the censorship of press freedom.

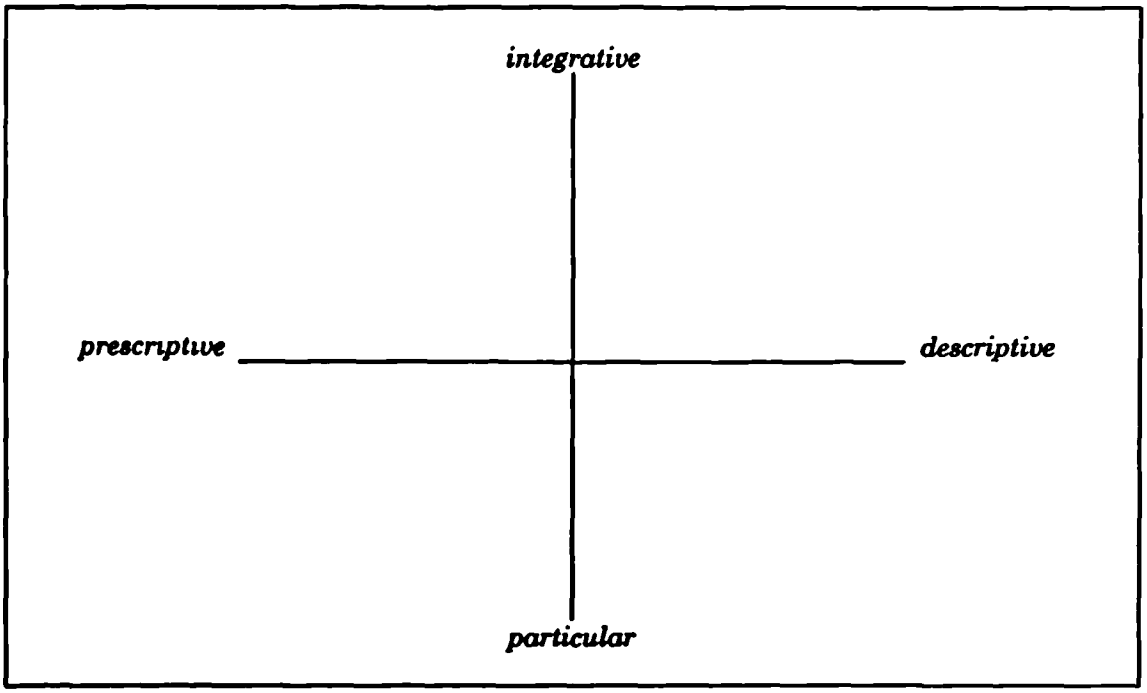
4.4 Tools for information policy research

Methodological problems exist in all disciplines; but information policy has more than its fair share. Despite all the difficulties and complexities surrounding information policy as an object of study, a considerable amount of material continues to be published on the subject. Whether this is an entirely healthy situation is open to question since, more often than not, its authors fail to make explicit either their methodology or their underlying assumptions.

Consideration of the methodological basis of information policy research is conspicuous by its absence in the serials literature. One of the few authors to tackle the issue is Trauth (1986) in a paper published nearly ten years ago which offers a highly generalised description of the information policy literature, locating studies in a two-dimensional matrix (see Figure 4.1 overleaf).

This matrix indicates the *scope* (vertical axis) and the *intent* (horizontal axis) of individual policy studies, regardless of their methodology. In this scheme, 'descriptive studies' are those which describe the current status of a policy or which highlight issues that should be addressed in its implementation. 'Prescriptive studies', on the other hand, are those which explicitly set out to make recommendations and to have an influence on the formulation of policy.

Figure 4.1: Information policy studies matrix



As well as differing in their intent, information policy studies have different scope characteristics. In Trauth's framework, the scope axis (particular-integrative) expresses both the breadth of coverage and the degree of interdisciplinarity reflected in a study; 'integrative studies' being those which go beyond the boundaries of a single discipline.

Content analysis frame 23: Intent of article

1	*Descriptive*
2	*Prescriptive*

Content analysis frame 24: Scope of article

1	*Particular*
2	*Integrative*

Trauth concluded that the biggest gap in the information policy literature (at the time of writing) was for studies falling into the top left-hand quadrant of the matrix: research which is both highly integrative and focused on the immediate needs of policy formulators. Published studies sharing both of these properties

are still rare; a function, perhaps, of the fact that they need, ideally, to be conducted on a large scale by multidisciplinary teams yet within the restricted time horizons of decision-makers. This section reviews some of the predominant modes of investigation which can be inferred from an examination of the literature; it is by no means comprehensive but I hope it will help us to draw some tentative conclusions for future, and possibly more productive, lines of enquiry. It is argued that five broad methodological strands can be inferred from a review of policy-related articles in the library and information science literature: approaches based on (1) classification; (2) the identification of policy issues and options; (3) reductionism; (4) forecasting and scenario-building; and (5) process-oriented research and case studies.

Content analysis frame 25: Information policy research strategies

1	*Classification*
2	*Issues and options*
3	*Reductionism*
4	*Scenarios and forecasts*
5	*Case studies*

Each approach serves different purposes and makes different sets of assumptions; and they are by no means mutually exclusive.

Table 4.3: Tools and methodologies for information policy research

Research tool	Methodological strengths	Methodological weaknesses
Classification-based approaches	• Useful tools for exploring patterns in complex data • Demonstrates breadth of issues embraced by information policy • Facilitates access to primary research materials	• Limited theoretical underpinnings • Obscures the political, social and institutional contexts within which policy operates • Allocation of policies into mutually exclusive taxa risks losing a sense of their interconnections

Table 4.3: Tools and methodologies for information policy research *continued*

Issue identification and options	• Primary value as a data collection tool • Useful for mapping and scoping complex policy problems	• Fails to provide an explicit framework for evaluating policy options • Typically generates highly value-laden results
Reductionism	• Reduces complexity and ambiguity to manageable proportions • Restricting analysis to a particular discipline (e.g. economics) allows underlying assumptions to be made more explicit	• May succeed in providing a cogent but partial explanation which is not useful in the real world *In extremis*, it becomes difficult to relate the parts to the whole
Scenarios and forecasts	• Generation of alternative visions is a useful input to decision-making • Reduces and constrains uncertainty	• Difficult to capture sufficient data to make valid extrapolations • Underlying models often too deterministic
Process-based approaches and case studies	• Highly integrative technique yielding 'context-rich pictures' • Useful for testing hypotheses and developing new theories	• Highly expensive of time and other resources • Difficult issues relating to access and confidentiality

4.4.1 Classification-based approaches

It is hardly surprising that faced with the enormous problems of scope and definition noted earlier, and given the natural predisposition of authors from the library and information science community, one methodological approach has been to identify and categorise the information policy literature. Bibliographies are not uncommon (Bennett, 1987; Julien, Robinson & Tinline, 1992; Dahlin, 1990) and they provide a very useful key to the growing body of official reports, policy guidance, statutes, books and journal articles. This bibliographic approach confers several advantages. It means that policy documents and studies are reasonably well documented and described and it facilitates access to a considerable amount of primary material for research purposes. As a tool for demonstrating the wide range of issues which fall under the general rubric of 'information policy', classification schemes fulfil a useful secondary role in drawing attention to the breadth of the subject.

The most traditional approach to categorising information policies and issues is to develop a classification which reflects *policy goals and objectives.* Chartrand (1986) and Milevski (1986) use the same scheme, which neatly reduces the scope of information policy to nine broad categories (Figure 4.2):

Figure 4.2: Chartrand's taxonomy of information policies

Government information resource management policy
Telecommunications and broadcasting policy
International communications policy
Information disclosure policy
Information, confidentiality and privacy
Computer regulation and computer crime
Intellectual property
Library and archives policy
Government information dissemination policy

Other authors, notably Hill (1994), derive hybrid literature-based classifications which may include a categorisation by institutional source (e.g. the European Commission), by document type (e.g. UK Acts of Parliament) and by policy goal (e.g. jobs and education). An alternative approach, adopted by Rowlands & Vogel (1991) blends a chronological description of policy events with goal-based subject access in the form of 'chronological trails'. These different approaches raise an interesting methodological question: to what extent is it possible to derive an understanding of information policy from a reading of the 'cold' policy literature as set out in White Papers and statutes or recorded in the form of administrative or court decisions? Information policy exists at two layers: that which is *explicit* and recorded in documentary form, and that which is expressed *implicitly* in the form of habits, received wisdoms, unwritten codes of behaviour, expectations and societal norms. As the influential US Rockefeller Report observes: "to debate whether there should be a national information policy is pointless ... There will be such a policy ... arrived at consciously or unconsciously, by commission or omission, carefully or haphazardly, in a comprehensive or piecemeal fashion" (SATCOM, 1969).

While there may be some practical benefits in a classification-based approach to information policy, there is little to recommend this approach from a theoretical

standpoint: classification can only deal with policy in a very superficial way, obscuring the political, social and institutional contexts within which policy is shaped and implemented. A more fundamental objection is that by classifying policies into mutually exclusive categories we risk losing a sense of the inter-relationships between groups of issues. Burger (1988) overcomes some of these objections by applying content analysis to the recommendations in the SATCOM Report, the result of a three-year long deliberation of the status and future requirements for scientific and technical information in the United States. In effect, Burger uses content analysis to systematically categorise and classify the not just the policy goals for each of the Report's recommendations, but also their rationale and the actors involved in their implementation.

In an interesting recent paper, Sillince (1994) has developed a *paradigm-based* (rather than a goal-based) categorisation as a means of demonstrating some of the ideological and institutional factors which he argues are responsible for the fragmentation of information policy at European Union level. Whether this represents a new lease of life for the classification-based approach is still questionable: perhaps classification is best regarded as a tool for the initial exploration and perception of pattern in complex policy data.

4.4.2 Issue identification and options

Without doubt, the most prevalent methodological position to be encountered in the information policy is the `issues and options' approach literature (Moore & Steele, 1991; McIntosh, 1990; Rowlands, 1995), although few authors are honest enough to make this explicit. This is an approach which seeks to identify the issues and concerns raised by the adoption (or potential adoption) of a specific policy, and then to suggest a range of possible actions to resolve potential conflicts and mitigate any foreseeable negative impacts. This style of research is often commissioned by policy-makers in government or industry as an input to the early stages of policy formation and is typically activated by a notable legal or public policy event (e.g. the publication of the Bangemann Action Plan or a technological advance or new application such as video-on-demand).

The function of 'issues and options' is primarily information-gathering. Its strengths lie in its currency and in its ability to map the policy terrain and scope the dimensions of a problem. Whereas classification-based approaches tend to focus on documents as the primary object of study, 'issues and options' typically draws on a much wider range of inputs, notably interviews and questionnaires directed at policy-makers and other stakeholders. This approach is therefore more integrative (in Trauth's sense) than the classification-based style of research, although it almost inevitably lacks the former's very explicit articulation of assumptions.

Examples of the 'issues and options' approach are pervasive in the journal and grey literatures. The methodological limitations are serious: most notably because the approach fails to provide a coherent framework within which policy-makers can properly evaluate the options open to them. There is a further danger that this style of research, while implicitly (or openly) appropriating values of independence and objectivity, is in fact methodologically vacuous and its results *sui generis;* neither robust nor amenable to verification. Furthermore, the results generated are often heavily value-laden. Despite these weaknesses, the issues-based approach emerges as the dominant methodological construct in the information policy literature.

4.4.3 Reductionist approaches

Earlier in this Chapter, reference was made to the normative basis of public policy. An argument was put forward which stressed the desirability of a more theoretically pluralist approach to information policy problems, drawing on multiple perspectives and traditions. As a call to arms to the policy studies community, this seems a reasonable proposition but it is a call to which few individual researchers can respond with any conviction or intellectual credibility.

The reductionist approach is one which, while (sometimes) acknowledging that a policy problem is inevitably multifaceted, seeks to reduce ambiguity by constraining data collection, analysis and interpretation within the framework of a specific discipline. Bushkin & Yurow (1981), for example, argue that "aspects

of United States domestic information policy can be divided into two broad categories: (1) the legal foundations of information dissemination and access; and (2) the economics and management of information". This framework is one which seems to have received the endorsement of a number of researchers who have adopted deliberately restrictive legal (Braman, 1988; Burkert, 1992) or economic (Porat, 1977; Greenstein, 1992) perspectives.

The reductionist approach is inevitable, given our need to make sense of extremely complex phenomena and, in many ways, it is highly desirable if it leads to a more rigorous consideration of information policy. Reductionism overcomes the vagueness of the 'issues and options' approach and enables underlying assumptions to be made much more explicit. Whether reductionist analysis leads to good policy, however, is a much more difficult proposition: given the complexity of the forces which drive and shape policy, to what extent can exclusively legal, economic or other single disciplinary models predetermine an overall 'optimal' policy regime? This is a fundamental problem of the reductionist approach: how can the parts be related to the whole? An analyst may get the organisational and economic aspects right, but ignore other factors that can be a decisive factor in the success or failure of a policy. Burger (1993) observes that there may be several reasons why information policy does not work. Specific policies may overlook contextual factors which impact on their implementation; the initial premises in the formulation of the policy may be incorrect; the policy may not be implemented as designed; or there may be unforeseen problems or situations about which no one has thought. Almost all of these reasons for a failed policy, however, have a common denominator of imperfect knowledge. Do policies fail because of ignorance or because society is too complex and things do not turn out as they were intended?

4.4.4 Scenario-based approaches

Forecasting studies and scenario-building represent another, quite different, methodological construct and one which is widely reported in the library and information science literature (Bates, 1990; Bezold & Olsen, 1986). A notable recent example of the species is the British Library's Delphi-style *UK*

Information 2000 study (Martyn, Vickers & Feeney, 1990). Scenario-building allows the construction of several alternative 'visions' of the future which policy-makers can use to make develop appropriate strategies for reacting to events. Scenario-building is a particularly attractive tool in the context of information policy problems because of the high levels of uncertainty which are inherent; especially those associated with the adoption and use of information technologies. High uncertainties can lead reasonable people to quite different judgements about information policy issues and the best ways of dealing with them. The central objective of scenario-building should therefore be to reduce uncertainty to manageable proportions. It is also one very effective way of accommodating the notion of stakeholders and the possibility of differential effects occurring in relation to various social and economic interest groups (Kiesler & Hinds, 1993).

Almost by definition, these futures-based approaches need to be highly inclusive. Burke (1994) notes that scenario-building is commonly based on the STEEP framework, which embraces Social, Technological, Economic, Ecological and Political perspectives. On the other hand, the limited effectiveness of forecasting and scenario development are often mercilessly exposed under the glare of hindsight. Two of the more obvious reasons for their comparative repeated failure are the extreme difficulty of capturing sufficient data from a broad enough base to make reliable extrapolations, and the use of conceptual models which are too deterministic, either in terms of the power of either technology or public policy to shape events.

Does the inevitable uncertainty which results from even the most meticulous forecasts imply a need for greater frankness by governments about the uncertainty of the environments in which they operate and their ability to deliver policies to precise targets? The study of forecasting raises questions of how decisions are made, not just an understanding of the techniques available.

4.4.5 Process-oriented approaches and case studies

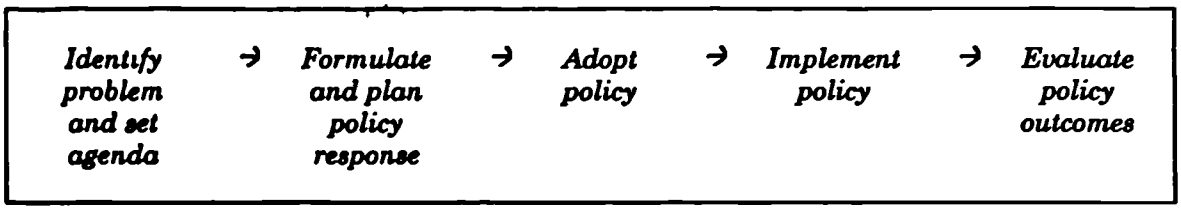
Despite the frequent assertion that information policy is enormously complex and multi-faceted, it can be argued that at least some of that 'complexity' derives from confusion and a lack of definition. Hogwood & Gunn (1984) note that there are many different uses of the word 'policy' in circulation. These include policy as a label for a broad field of activity (a government's overall 'economic' or 'industrial' policy); as a specific proposal or decision of government; as a programme or legislative outcome; as a theory or model; or as a *process*. Conceptualising public policy as a process, rather than as an outcome or an event, is the dominant theoretical approach in the general policy studies literature. The reasons for this emphasis are not too difficult to understand: it enables us to understand how the complexities of the policy-making process unfold over time and thus raise some fundamental questions, such as who are the 'policy-makers' and how do institutional factors bias policy-making?

One useful approach to understanding information policy is to employ a systems approach. This recognises that policy-making comprises a series of inputs (people, ideology, expediency, information, research, investment) and outputs (wealth creation, better healthcare, access to democracy). By conceiving of information policy-making as an INPUT-PROCESS-OUTPUT (I-P-O) model for organising our thinking, we shift to the view that information policy governs a *process* (such as the storage and transmission of information) rather than a *thing* (such as technology). 'Information policy' might therefore be better thought of as a verb rather than a noun. Just as the I-P-O model can be used to describe how data are transformed into information and then knowledge, so it can also offer insights into policy-making. So, rather than addressing policy issues relating to a specific advance in software or data communications (technology-driven) we can switch our attention to the underlying functional aims and objectives of policy (Trauth, 1986).

Conceptualising policy as a *process*, rather than a specific outcome or event, is very useful. It helps us to understand how policy develops over time and how policy is shaped by (and, in turn, shapes) organisational and social factors.

Policy is not an abstract ideal, it takes place in an imperfect and sometimes confusing world. A typical representation of the policy-making process is the 'functional staged model' (Lasswell, 1970). A simplified version is illustrated below (Figure 4.3).

Figure 4.3: Staged model of the policy-making process



As we move from left to right, a problem is first recognised and defined and then finds itself on a decision-maker's agenda. Alternative 'solutions' are developed, presented and rejected in favour of the option which offers the maximum net benefit (or is the most convenient, expedient or cheapest). This is then officially adopted. Implementation begins and some kind of evaluation or monitoring procedures are usually invoked so that any undesirable outcomes can be identified and mitigated against. In many cases, the results of that evaluation will require adjustments to be made earlier in the chain, perhaps resulting in a complete re-design of the policy.

To a large extent, once a policy-making process gets underway, it tends to be continuous. It has been said that policy-making has "no beginning and no end" (Lindblom, 1980). This overstates the case, as it is possible to define reasonable starting and termination points in a pragmatic way. Within those bounds, however, the process can be regarded as continuous.

The power of the staged model is that it offers us a way of grouping a wide range of apparently disconnected decisions, phenomena, observations and data into meaningful units. It also has a certain intuitive appeal which is not easy to discount. It also has its limitations. Many critics would immediately point out that real life, with its rough-and-tumble of politics and horse trading, is not nearly as tidy as the model suggests.

Nonetheless this is how policy-making is most often presented in the media and many policy-makers would still justify and defend their own actions, however apparently irrational at the time, in terms of this ideal framework. Clearly theoretical models can be very powerful and may shape the course of events in the real world. Quite how different theoretical models interact and impact on the course of events has yet to be examined in the context of information policy studies but it might well prove a productive *research topic.*

Content analysis frame 26: Stage in policy life-cycle

1	*Policy design*
2	*Policy implementation*
3	*Policy evaluation*
4	*Multi-stage*

From the point of view of our search for methodological tools and approaches to information policy problems, this framework appears to offer a number of potential benefits: it represents an antidote to the narrowly reductionist approach by allowing us the possibility to build 'rich pictures' of information policy and its evolution. There are few examples of a process-driven analytical approach to information policy in the literature, however. One notable exception is Karni who, in a paper on the formulation of information policy within organisations (Karni, 1983), suggests that a framework for understanding policy formation should comprise four elements: information policy objectives; measures to determine how well these are achieved; policy actions and alternatives; and a consideration of the information services system itself.

Closely related to the process-oriented approaches and partially overlapping is research based on case studies. Several detailed case studies (Burger, 1988; Dawes, 1991; Jacobson, 1989; Stewart, 1990) of the history and evolution of US public policies are available, and they demonstrate the importance of institutional, historical, cultural, political and human factors in determining information policy, throwing further doubt as to the usefulness of the narrow reductionist approach.

The particular value of case studies is that it they can reveal much about the policy-making process that might otherwise be hidden from historians or other researchers relying primarily on printed documents. Guiding philosophies about government and information are not the only determinants in the policy debate; policy formation and implementation may also be shaped by staff in the legislative and executive branches of government. A major limitation of case studies is that they do not permit generalisation. They can be used, however, to test existing theories; to develop new theories; to provide detailed contextual analysis of particular events; and to help provide an 'intuitive feel' for the subtleties and nuances of the policy process.

4.5 Conclusions

The picture which emerges from this brief review of the methodological tools employed by writers on information policy is not a flattering one. Compaigne (1988) notes that the information policy debate is highly political and that whether we are discussing issues such as ownership, copyright, equity, or literacy, the 'right' answer is (and should always be) subject to legitimate debate. There is considerable value, therefore, in reminding ourselves that a truly open and democratic debate should make explicit the variables which go into determining the 'right' answer. What assumptions are being used? Is there another set of reasonable assumptions that would yield different answers? What is the agenda of the policy-makers? Do they state their agenda explicitly, or is it hidden?

What is needed and what does not yet appear to exist is a body of knowledge and research tools that can provide a *value-critical* and *paradigm-critical* approach to the study of information policy. The methodologies for policy analysis are themselves rarely 'value-free'. Many techniques have unstated but built-in assumptions about values. This means that, in some cases, a particular technique may be chosen for 'political' purposes, to advance a cause rather than in a genuine spirit of scientific enquiry.

This Chapter deliberately does not conclude with a blueprint for undertaking information policy research; we are a very long way yet from developing a robust box of tools and methodologies. Nonetheless, the limitations of much of the current reporting in the library and information science literature suggest some desiderata for research design in information policy:

1. Greater recognition of the need for a more *value-critical approach* to information policy problems.

2. The need to distinguish more clearly between *descriptive* and *prescriptive* modes of policy analysis: between what policy *is* and what it *ought* to be.

3. Greater awareness of the dangers of confusing the treatment of information policy as a *dependent* or an *independent* variable in analysis.

4. The need for a broader range of *more sophisticated indicator data* on the social, cultural and economic impacts of information policy.

5. Greater clarity in terms of describing the *scope* of information studies and the *definition* of terms used.

6. The need to make assumptions, methodologies and working hypotheses *more explicit*.

7. Greater awareness of the *hierarchical* nature of public policies ('no policy' does not necessarily mean no policy).

8. A more imaginative and *catholic* approach to the appropriation and testing of research tools from other policy disciplines.

9. A more creative approach to information policy analysis and a greater willingness to develop *explanatory models*.

10. The need to retain a *multidisciplinary integrative perspective,* even in small-scale studies.

11. Greater recognition of the *discipline-bounded* and *technology-driven* nature of much information policy research.

12. Greater awareness of *context* and the factors which influence the formulation, adoption and impact of information policies.

Part Two
Experimental evidence

Chapter 5: The research design

> *"A sea change in bibliometric research has occurred, but as yet it is fully understood by only a few hundred researchers world-wide. Each year in the 1990s will be marked by new waves of studies, larger, better designed, and better analysed than those of previous decades"*[15]

5.1 Introduction

This section of the thesis reports on a series of empirical investigations into the structure of the information policy serials literature.

The central resource for these studies was a bibliography specifically created for the purpose (see Appendix B). The mechanics of the creation of this resource is fully described in Section 5.3; for the moment, it is sufficient to note that the test collection represents a longitudinal sample of 771 serial articles published during the period 1971-1996 and drawn from the multidisciplinary *Social Science Citation Index*.

The research findings presented in the four Chapters which follow are based on extensive secondary data analysis, a point which itself raises an important methodological issue. The questions asked in bibliometric research are more contextual than the variables which are studied; these tend to be fairly uniform across the whole field of scholarly communication for two reasons: (a) the consistent properties of text, and (b) the conventions of secondary publishers like *ISI*. For this reason, some care has been taken to contextualise the experimental findings, in so far as this is possible, by reference to other published work and the author's knowledge of the field. It is hoped that by approaching the same data set with different techniques and assumptions, that at least some of the

[15] Paisley (1990:299).

evident weaknesses of each method will be brought under a measure of control[16]. These issues are discussed more fully in the concluding Chapter.

The techniques which were used to mine useful indicators from the document test collection are well-established and understood: they include a bibliometric census study, an exercise in hierarchical document clustering, an author cocitation study, and content analysis. In each case, unobtrusive measures were produced, measures which did not influence the phenomena being described.

The overall research design, and the selection of the four techniques mentioned above, was partly driven by practical considerations, partly by methodological opportunism, and partly as the result of reading a stimulating article on future directions in bibliometric research by William Paisley. Paisley argues that three generations of bibliometric research can be identified: the first two based respectively on text-based and citation-based measures. The third (future) generation of studies will, in the author's words be marked by "the complementary use of both approaches, combined with an increasing use of supplementary measures from non-bibliometric sources" (Paisley, 1990:281).

As it happens, Paisley's article was not discovered until about half of my empirical research had already been completed. In the concluding pages of his article, Paisley sets out a programmatic description of what questions a third-generation bibliometric study might address. I quote from this article at some length, since many of the issues raised were in fact already evident in my research design:

> *"Drawing upon the parallel with indicators research, we note a missing level of questions on the fundamental side of those raised by Borgman*

[16] Webb and others (1966:1) emphasise "the necessity for multiple operationalism, a collection of methods combined to avoid sharing the same weaknesses". They discuss two categories of weaknesses: lack of internal validity (uncertainty in explaining the causes of the observed differences in the sample itself) and lack of external validity (uncertainty in generalising the observed difference to other samples).

> *[Borgman, 1990]. These are 'demographic' questions about the publications system and the social system behind it, such as:*
> *(a) numbers of articles published in scholarly subfields;*
> *(b) numbers of articles published on a wide range of important topics;*
> *(c) number of researchers actively publishing in each subfield according to the tabulations in (a) above;*
> *(d) number of researchers actively publishing on each topic according to the tabulations in (b) above;*
> *(e) national or geographic location and institutional affiliation of researchers identified in these analyses;*
> *(f) patterns of team research implied by the authorships per subfield or topic; or*
> *(g) characteristics of published articles per subfield or topic, including conceptual structures in the text, number of citations, average age of journal citations, and so on.*
> *In addition to the value of these demographic findings in themselves, they provide parameters for interpreting research on higher-level questions about social networks, importance and diffusion"* (Paisley, 1990:297).

Of the questions raised above, (b) and (d) through (g) are tackled explicitly in this thesis. Questions (a) and (c) are addressed, but without extensive reference to neighbouring scholarly subfields (surely Paisley's intention), such as information law, information management or public administration.

5.2 Overview of Chapters 6-9

5.2.1 Chapter 6: Simple bibliometrics

This Chapter examines the distribution of various bibliometric elements across the test collection and through time, using univariate or bivariate analyses. It examines patterns of growth, knowledge accumulation, ageing and obsolesence, authorship and Bradford scattering.

5.2.2 Chapter 7: Cluster analysis

This Chapter employs hierarchical clustering techniques to investigate any underlying natural structure in the document test collection, based on classifications generated from a range of bibliometric and content-based indicators. The Chapter essentially addresses the question of whether the bibliography is best considered as a single homogenous entity, or as a series of joint bibliographies distinguishable by some, initially unknown, criterion or criteria.

5.2.3 Chapter 8: Author cocitation analysis

Author cocitation analysis is a methodology which has been claimed to offer unique insights into the social and intellectual structure of a field of study. This Chapter explores the patterning evident in the cocitation patterns of 21 leading information policy authors using a battery of multivariate statistical techniques. It provides an example of what Marshakova-Shaikevich (1993) terms 'structural' bibliometrics; unlike the approach in Chapter 6, the emphasis here is on the connections between objects rather than on 'simple' distributions. The results are corroborated by means of a simple postal questionnaire.

5.2.4 Chapter 9: Content analysis

The simple bibliometric analysis presented in Chapter 6 was based on elements which were either already present in the records downloaded from the *Social Science Citation Index*, or which could be directly derived from them. The content analysis extends this earlier analysis by investigating the distribution of various content indicators, inferred from the original full-text documents. These indicators are of a highly qualitative nature and seek to describe and codify such factors as the aims, scope and intentions of published articles; the methodologies used; and the stage of the policy cycle being investigated.

5.3 The document test collection

5.3.1 The data source: Social Science Citation Index

There are a number of abstracting and indexing services which covers subsets of the information policy literature: *Public Affairs Information Service, Library & Information Science Abstracts, Information Science Abstracts* and *Applied Social Science Abstracts* to name the most obvious. A comprehensive bibliography in the area of information policy would need to investigate all of these sources as well as published bibliographies by authors such as Hill (1995).

In common with the majority of bibliometric studies, the work reported here is based entirely on a single data source: the *Social Science Citation Index*, a decision which requires some justification. While each of the secondary sources mentioned above may be expected to contain relevant articles, they are highly field-specific in the sense that they index comprehensively within a narrowly prescribed list of titles. Since one of the aims of this thesis is to explore the multidisciplinary nature of information policy research, the use of these tools would introduce an unacceptable degree of distortion. Rather than seeking a high level of comprehensiveness, the intention behind the compilation of the test collection was to seek reasonably balanced coverage across contributing literatures. Other pragmatic justifications for choosing *SSCI* above other sources may be summarised as follows:

- *ISI* editorial policy emphasises high-quality (usually peer-reviewed) journals with high international impact;

- Depth of time coverage: in its *DataStar* implementation, *SSCI* records are available over a quarter of a century (1971 to the present) making it possible to analyse changes in the evolution of the information policy serials literature;

- The *SSCI* record structure contains elements which are otherwise unavailable without recourse to the source documents. The most notable

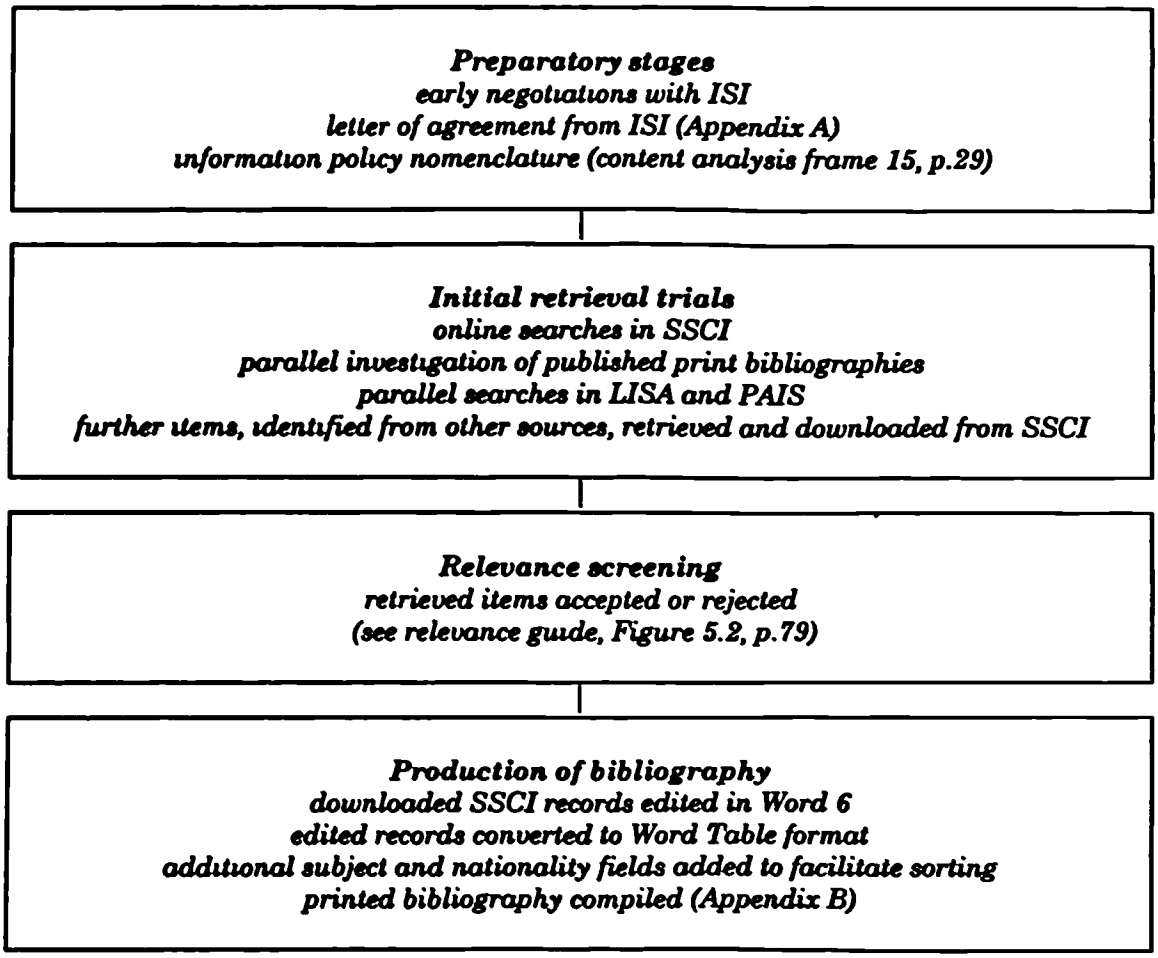
elements are journal category, corporate addresses and details of citations;

- The database is available free at the point of use to members of the academic community (via *BIDS*).

5.3.2 Preparatory stages

The preparation of the document test collection took place in four phases which are represented in Figure 5.1:

Figure 5.1: Compiling the document test collection



Negotiations with the Institute for Scientific Information

It is a strict condition of the use of *ISI* databases that the data cannot be used for the purposes of bibliometric study without written permission. I am very grateful to Robert Kimberley, Managing Director, European Branch for granting me permission to use the data in this way (see Appendix A).

Apart from dial-up online access through academic and commercial hosts, *ISI* offer a number of services to the research community, including the possibility of ordering customised data sets. The advantages of buying a tailored data set, rather than searching public online services, is that additional fields are available: these include indexing by second and subsequently named authors, and richer citation information than is available in the various online implementations.

In the early stages of planning the experimental work, the possibility of obtaining a data set directly from *ISI* was explored. The main concern was to obtain the data in a format which would facilitate author-author and journal-journal citation analyses. Apart from the expense involved (which would have been in four figures!), an even more intractable problem was encountered. Many of the users of *ISI's* customised research services are interested in a particular journal, journal category, or institution. As will be seen later, the search strategy used to recover articles on information policy was very heuristic and complex. It was not possible to operationalise the search in a form that *ISI* could use, other than a list of *DataStar* accession numbers, which, as it turned out, *ISI* were anyway unable to map onto their database. The regrettable outcome was that citation analysis, whose application in the context of this thesis might have been very powerful, became totally impractical.

5.3.3 Initial retrieval trials

The search strategy

Formulating a search strategy which is broadly representative of the information policy construct is not a trivial task. Matters are made more difficult by the lack of an active indexing policy by the compilers of the *SSCI* database. Keywords and subject descriptors are only available if and when supplied by the author in the original article: no attempt is made by *ISI,* for perfectly understandable reasons, to offer a common indexing platform across the whole database.

The basic search strategy adopted was therefore an heuristic one. Search terms and formulations were derived from the subject classification scheme (Content analysis frame 15, overleaf) and the initial results inspected for author-supplied descriptors and keywords which might prove to be useful in further searches[17]. The bulk of the searching was conducted using the *BIDS* implementation of *SSCI,* although considerable use was also made of the *DataStar* implementation. Two major drawbacks were noted in relation to *BIDS:* records are available back to 1981 only (1971 in the case of *SSCI* on *DataStar*) and the search features and user interface are relatively limited. The research took place during June and July 1996, which means of course that the records for 1995 are almost certainly incomplete.

Maximising recall

Online searching was continued in this highly interactive manner until a point was reached at which relatively few new records were being identified. At this point, other techniques were employed in an attempt to gauge and maximise the level of recall being achieved.

The first test of recall was to compare the online search results obtained so far with three formal published bibliographies on information policy by Rowlands & Vogel (1991), Julien, Robinson & Tinline (1992) and Hill (1994). This yielded a

[17] The phrase 'information policy' occurred in only 124 of the 771 records.

relatively small number (38) of references which had been missed in the online searches but which were subsequently found to be present in the *SSCI* database.

Content analysis frame 15: Subject analysis of article

10	***General articles on information policy***
11	*Theoretical aspects of information policy*
12	*National and international information policies*
20	***Information infrastructure policies***
21	*Research & development (including STM information policies)*
22	*Libraries, archives and public records*
23	*Telecommunications, broadcasting and information superhighway policies*
24	*Information technology policies*
30	***Information management in government***
31	*The collection and acquisition of information resources by government*
32	*Information resource management in government: policies and practice*
33	*Government information systems, clearinghouses and dissemination policy*
40	***Information access and control***
41	*Freedom of access to information*
42	*Confidentiality and personal privacy*
43	*Information control on grounds of national security*
50	***Information industry policies***
51	*Information standards and protocols*
52	*Copyright, intellectual property and information law*
53	*Regulation of the information industry and information markets*
54	*Trade in information services and transborder data flows*
55	*Public-private relationships in the information industry*

A second test of recall was made by means of a highly selective search of *Library and Information Science Abstracts (LISA)* on *DIALOG, Science Citation Index* on *BIDS* and the CDROM version of *Public Affairs Information Service (PAIS)*. This consisted of searching free-text for occurrences of the phrase 'information policy' in records referring to serial articles published during 1985, 1990 and 1995. This approach yielded only eleven relevant articles (eight from *LISA*, two from *SCI*, one from *PAIS*) which had been missed in the original online searches of *SSCI* and a decision was taken to draw a line under any further research.

Relevance screening

The broad nature of the subject being investigated and the absence of a common indexing platform in the *SSCI* database meant that a great deal of material which was judged to be irrelevant was recovered during the online searches. As a guide to relevance judgements, the following scheme of inclusion and exclusion criteria was developed:

Figure 5.2: Test collection inclusion and exclusion criteria

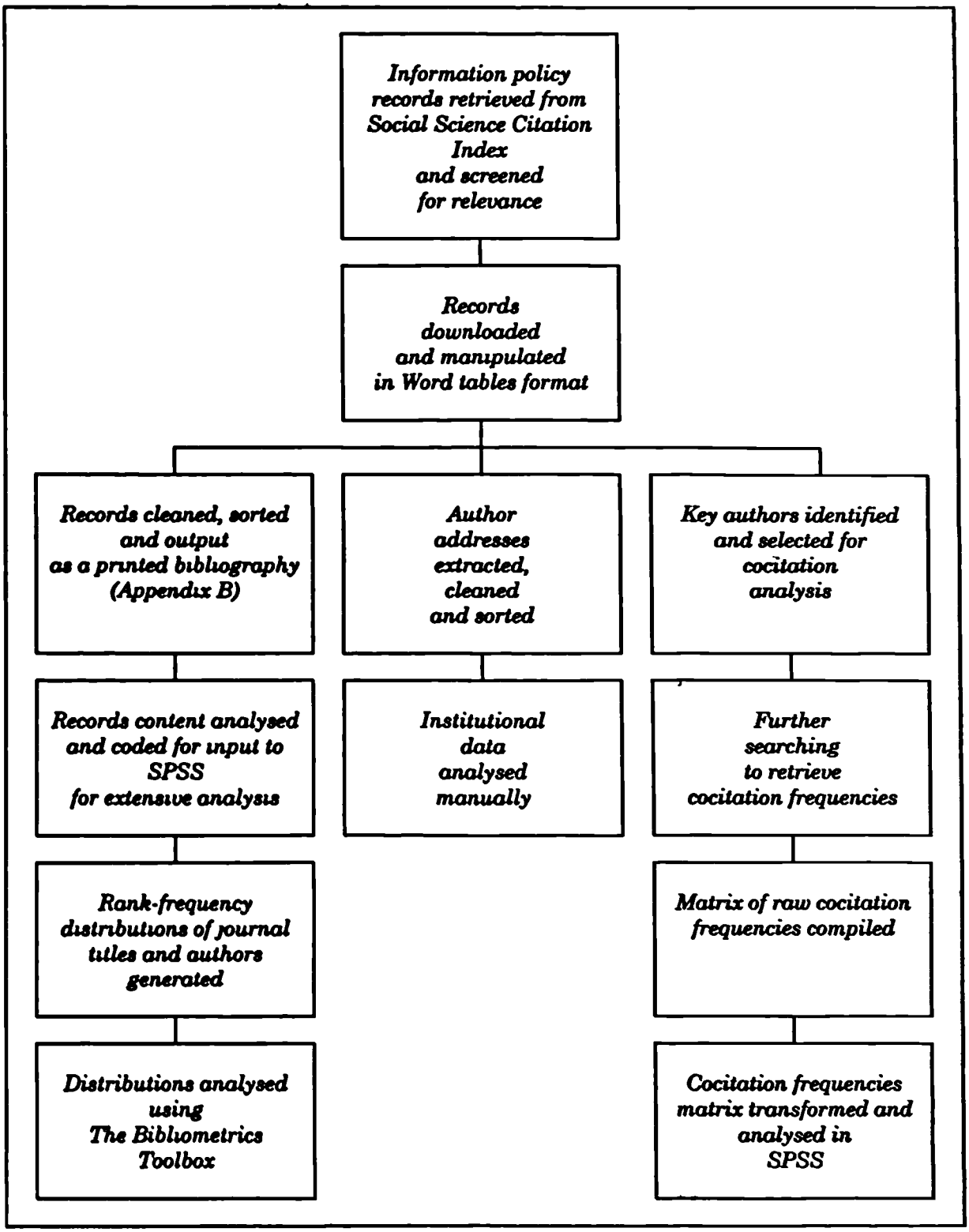
Included:	*Excluded:*
Papers *Notes* *Review articles*	*Bibliographies* *Book reviews* *Letters* *Editorials* *Abstracts* *Discussions*
Three or more pages in length	*Shorter than three pages*
Articles dealing with issues of general public policy	*Articles dealing with issues of narrow organisational interest (e.g. the acquisitions policy of a government agency library)*

The application of these criteria resulted in the rejection of nearly 400 'hits' from the *SSCI* database. The final bibliography consists of 771 serial articles and exhibits a reasonable spread of coverage against each of the main and sub-categories of the faceted subject classification. A further decision was taken to restrict the bibliography to journal articles. Journal articles represent contributions to the literature that have been subjected to review by the authors' peers and found acceptable for publication whereas other forms of publication such as reports, conference papers and book reviews are often subject to little or no review.

5.4 Data preparation and analysis

5.4.1 Production of the printed bibliography

Figure 5.3: Overview of data preparation and analysis



The downloaded *SSCI* records were manipulated in Word for Windows version 6.0, using the sorting capabilities offered by the Tables feature for the production of the final bibliography. In a bibliography covering such a long period of time, it was not surprising to find that a number of journals had changed their title (for example, *Government Publications Review* became the *Journal of Government Information* in 1993). A comprehensive check of Ulrich's directory was therefore made to ensure that all instances of journal title changes were picked up. In the analyses which follows, articles are attributed to the *most current* journal title, irrespective of their date of publication. The final bibliography is included as Appendix B to this thesis.

5.4.2 The Bibliometric Toolbox

Much of the analysis presented in the next Chapter was facilitated by the use of *The Bibliometric Toolbox* (version 2.8), a freeware package[18] developed by Terry Brooks at the University of Washington. This software allows for two forms of data input. Downloaded records may be processed directly using a pre-editor in a way that permits both the production of a bibliography and its subsequent bibliometric analysis. Alternatively, the software performs a range of bibliometric analyses on rank-frequency distributions. This latter form of indirect data input was used for the purposes of the analysis here.

The Bibliometric Toolbox also provides some sample data which make it possible to emulate Bradford's work on the lubrication literature and Goffman and Warren's work with the mast cell literature. Manual calculations of a number of bibliometric measures have confirmed the reliability and accuracy of *The Toolbox*[19].

[18] Copies of the software are available by anonymous ftp at the following location: *ftp.u.washington.edu/public/tabrooks/toolbox*.
[19] See review article by McLain (1990).

5.4.3 Institutional affiliations file

The printed bibliography does not include institutional affiliations. These elements of the downloaded records were extracted and stored separately as a Word for Windows table. The initial list of 'dirty' corporate affiliations required a great deal of manual effort, first of all to sort out corporate name variants, and then to decide upon an appropriate level of aggregation. Corporate affiliations were aggregated at the highest reasonable level possible: for example, the Department of Media Studies, the Graduate Program in Mass Communication, and the Faculty of Communication at the University of Texas were combined into a single, university-wide, affiliation. Where only one department or research team was retrieved, that narrower corporate affiliation was used.

Finally, each institution was then assigned to one of five sectors (Education, Government, Non-Governmental Organisations, Non-profit, Industry) for the purposes of describing the dispersion of knowledge production in information policy and exploring patterns of author collaboration.

5.4.4 Statistical Package for the Social Sciences

Much of the analysis presented in the following Chapters was derived using the *Statistical Package for the Social Sciences (SPSS)* version 7.5 (for Windows 95 / Windows NT), with considerable use being made of the *Professional Statistics™* module[20]. A data grid was created where each row represented an article and each column a primary or derived variable. A detailed description of all these variables is attached as Appendix E.

[20] See Norušis (1994) for a technical description of the *Professional Statistics™* module.

Chapter 6: Simple bibliometric analysis

> *"Despite its abstract air and disembodied rhetoric, bibliometrics is grounded in the patterned behaviour of human beings"*[21]

6.1 Introduction

In studies of the natural and social sciences, much attention has focused on differences in cognitive structure and patterns of communication in various disciplines. Since Price's work on communication habits in the hard and soft sciences (Price, 1970), an increasing number of sociologists of science have been turning to bibliometric techniques to investigate these questions.

Two broad methodological directions have been observed within the field of social bibliometrics by Marshakova-Shaikevich (1993) which she distinguishes as 'simple' and 'structural' bibliometrics. In simple bibliometrics (sometimes called census studies) the researcher is primarily concerned with the distribution of various bibliographic elements (authors, subject headings, dates of publication) *across* a collection of documents. In structural bibliometrics, on the other hand, the emphasis is on the connections and linkages *between* objects, in their correlation and classification.

This distinction is mirrored in the structure of this thesis. This Chapter offers an overview of the characteristics of the information policy test collection in simple bibliometric terms. The bibliographic elements studied here are restricted to those which *ISI* makes available in its online implementation of the *Social Science Citation Index*. This simple bibliometric approach is extended in Chapter 9, where a richer set of bibliometric elements is investigated by means of content analysis. The basic approach is the same, to paint a picture of the structure and dynamics of the information policy literature, albeit this time from a richer palette. Chapter 8 is a study of prominent authors in information policy

[21] White & McCain (1989:123).

in terms of the patterning evident from the way that they have been cocited by others. In Marshakova-Shaikevich's scheme, this study clearly sits within the structuralist camp, since it explores linkages between objects, in this case authors.

The quotation at the top of the previous page serves as a reminder that bibliometric indicators are not an end in themselves; they are simply tools which help to construct a better understanding of the intellectual and social structure of information policy studies.

6.2 Research questions

Table 6.1: Research questions and related bibliometric indicators

Research questions	Bibliometric indicators	Refer
Summary characteristics		
How many authors are represented in the bibliography?	Total number of individual personal and unique corporate authors	*Fig.6.1*
How many journals carry information policy articles?	Total number of unique journal titles	*Fig.6.1*
Which languages are represented?	Distribution of articles by language	*Fig.6.2*
Which countries are represented?	Distribution of articles by geographical provenance and growth over time	*Fig.6.3* *Fig.6.4*
Where do authors publish?	Distribution of articles by *ISI* journal category	*Fig.6.5*
Patterns of growth		
Is the production of information policy articles increasing?	Article production in two-year increments	*Fig.6.6*
If so, what is the pattern of growth?	Regression model fitted to cumulative article growth over time	*Fig.6.7*
Is information policy growing relative to the social sciences serials literature as a whole?	Growth of information policy articles expressed as a proportion of all articles added to *Social SciSearch*	*Fig.6.8*

Table 6.1: Research questions *continued*

Knowledge accumulation		
How 'scholarly' are the articles in the bibliography?	Ratio of articles without citation lists to those with (the 'Windsor ratio')	*Fig.6.9*
Is the literature becoming more or less 'scholarly'?	Changes in the 'Windsor ratio' over time and balance of research / opinion articles	*Fig.6.10* *Fig.6.11*
Is the information policy literature 'ephemeral' or 'classical'?	Average value and frequency distribution of Price's index	*Fig.6.12*
Is the value for Price's index typical of the social sciences?	Comparison of Price's index values with other studies	*Fig.6.13*
Is this the observed value of Price's index a structural feature?	Changes in the average value of Price's index over time	*Fig.6.14*
Ageing and obsolescence		
What is the pattern of citation maturation and decline?	Distribution of citations to articles over time	*Fig.6.15*
What is the half-life of a typical citation?	Regression model fitted to above	*Fig.6.15*
Bradford scattering		
Is the bibliography evidently incomplete?	Evidence of a Groos droop in the Bradford bibliograph	*Fig.6.16*
Does the Bradford bibliograph conform to the classical J-shaped curve or does it show the concavity sometimes encountered in the social sciences?	Deviation of the Bradford bibliograph from linearity	*Fig.6.16*
How scattered is the bibliography?	Bradford multiplier, b_m	*Tab.6.3* *Tab.6.4*
How strongly clustered is the bibliography?	Clustering index Average recurring productivity	*Fig.6.17*
How evenly spread are the articles across Bradford zones?	Egghe's C prime index	*Fig.6.17*
Are the core journals library and information science titles?	Cross-tabulation of journal titles by Bradford zone and journal category	*Tab.6.5*
How homogenous is the information policy bibliography?	Comparison of article/journal density with other studies	*Tab.6.13*
Patterns of authorship		
Who are the most productive authors represented in the bibliography?	Top ranking personal authors by article production	*Tab.6.6*
How productive are information policy authors?	Calculation of exponent n by fitting a linear regression model to the Lotkan distribution	*Tab.6.7* *Fig.6.18*

Table 6.1: Research questions *continued*

Is author productivity typical of the social sciences?	Comparison of exponent *n* with other published studies	*Tab.6.8*
How extensively do information policy authors collaborate?	Distribution of articles by number of joint authors	*Tab.6.9*
Is the incidence of author collaboration increasing or decreasing?	Mean articles per author over time	*Tab.6.10*
Where are information policy corporate authors located?	Distribution of corporate authors, globally and by EU countries	*Fig.6.19* *Fig.6.20*
Are most information policy corporate authors based in the university sector?	Distribution of corporate authors by sector	*Fig.6.21*
Which are the most highly productive corporate information policy authors?	Ranked listing of highly productive corporate authors	*Tab.6.11*
How concentrated or dispersed is the institutional basis of information policy research?	Tabulated Herfindahl indices by sector and over time	*Tab.6.12*
What is the pattern of research collaboration at institutional level?	Incidence of collaboration within same institution, within same sector, across different sectors	*Fig.6.22*

6.3 Distributions of articles

This Chapter presents some of the key bibliometric characteristics of the document test collection, following the general approach developed by Donohoe (1973). Wherever possible, the reported bibliometric findings are related to other published studies, drawn mostly from the library and information science literature. This is intended to locate the results within a more meaningful context, although the reader is cautioned from drawing *direct* comparisons between this and other bibliometric studies. In a critical review of bibliometric and other science indicators, King (1987) notes that there are no commonly-agreed standards or methodological guidelines for conducting bibliometric studies and so there are few, if any, reliable field-independent indicators.

6.3.1 Summary characteristics of the test collection

The purpose of this Chapter is to provide a general picture of the information policy serials literature which may also be of more general interest to researchers and information specialists. More specifically, the research

objectives address the questions posed in Table 6.1 by examining how various elements of the *Social Science Citation Index* record structure are distributed across the test collection. For convenience, the document test collection is referred to throughout as the *main corpus* and denoted by *M*.

Figure 6.1: Main corpus: summary bibliometric indicators

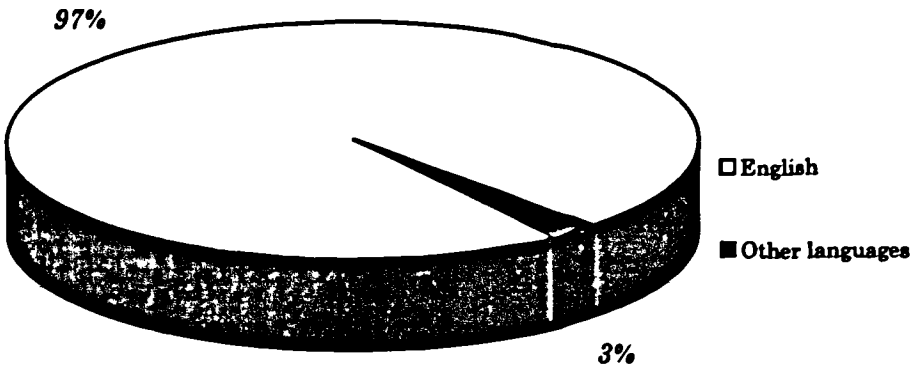
Primary indicators:	
Time period	*1972 to mid-1996*
Number of articles, A	*771*
Number of serial titles, J	*181*
Number of authors, a	*632*
Number of corporate authors, c	*231*
Number of works cited in A	*16,229*
Number of citations to M corpus articles in Social SciSearch database	*1,191*
Derived indicators:	
Articles per journal title (A/J)	*4.26*
Articles per author (A/a)	*1.22*
Articles per corporate author (A_c/c)	*2.77*
Authors per article (a/A)	*1.03*

The distribution of articles by language

As expected, English is by far the most frequent language represented in the bibliography (see Figure 6.2 overleaf), accounting for 97 per cent of all articles. The other languages represented are German, French and Russian[22].

[22] In a 1983 study of five abstracting & indexing services covering aspects of the library and information science literature *(Library & Information Science Abstracts, Information Science Abstracts, Computer & Control Abstracts, Referativnyi Zhurnal Informatika,* and *Bulletin Signaletique),* Bottle & Efthimiadis (1984) found that the overall proportion of English language materials was 71.1 per cent. The article claims that similar patterns of distribution of the English language obtain in the social and natural sciences.

Figure 6.2: Distribution of articles by language



This finding raises a fundamental methodological concern—the extent to which the distributions observed in the test collection are generalisable to the information policy serials literature as a whole. There is much discussion in the bibliometrics literature concerning the incompleteness of *ISI* database files and the issues that this raises[23] (see, for example, Osareh, 1996b:221). The limitations imposed by *ISI* editorial policy (and the possibility that this policy may have changed over time) are difficult to avoid: Kārki cautions that when *ISI* citation indexes are used as a source, the resulting coverage "can be nothing but biased in favour of Anglo-American research" (Kārki, 1996:329). Where possible, and especially in Chapter 9 which reports on a content analysis of the test collection, the findings are expressed in terms of *relative* rather than absolute change. Similar considerations apply to the geographical and disciplinary distributions of articles illustrated in Figures 6.3 and 6.5. King further points out that publication practices vary across fields and between journals, and that the 'pressure to publish' is influenced by institutional factors such as the emphasis placed on publication counts to secure tenure, promotion, or research grants (King, 1987). While acknowledging that publication counts are not in any

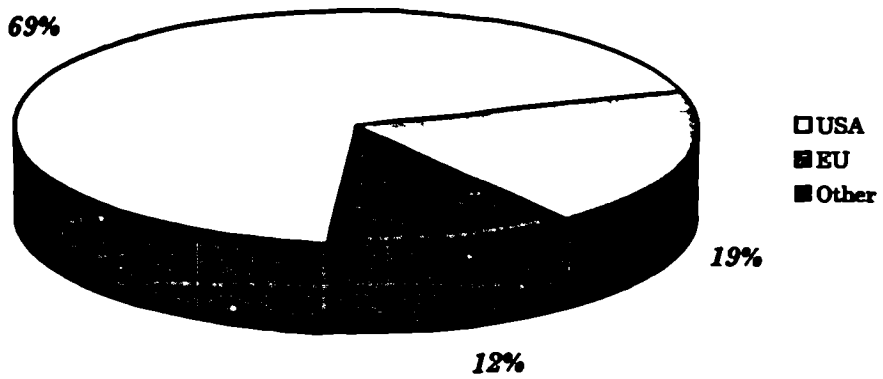
[23] In a systematic search for articles containing the phrase 'Information Society' in four abstracting & indexing services *(Information Science Abstracts, Social SciSearch, INSPEC* and *Arts & Humanities Search)* Duff (1995) found that *Social SciSearch* contributed 58 per cent of all the references recovered. *INSPEC* yielded a slightly higher proportion of unique references than *Social SciSearch*, from which Dunn concludes that "it would appear that the social implications of developments in computing and telecommunications technologies are now being contemplated, not only by information professionals and social scientists, but also by researchers in the so-called hard sciences".

way an accurate reflection of quality, they may still provide a crude measure of research output.

The distribution of articles by geographical provenance

The geographical provenance of articles in *M* (as determined by the corporate address of the first author) shows that US contributions figure strongly in the test collection. This is unsurprising, and not only because of the purported Anglo-American bias of *ISI* editorial decisions: the USA has a sophisticated and highly developed set of public information policies and possibly the most highly institutionalised environment for information policy research (see Table 6.11, p.120). The Canadian contribution (4.3 per cent) further illustrates the extent to which North American writers are highly represented in the bibliography.

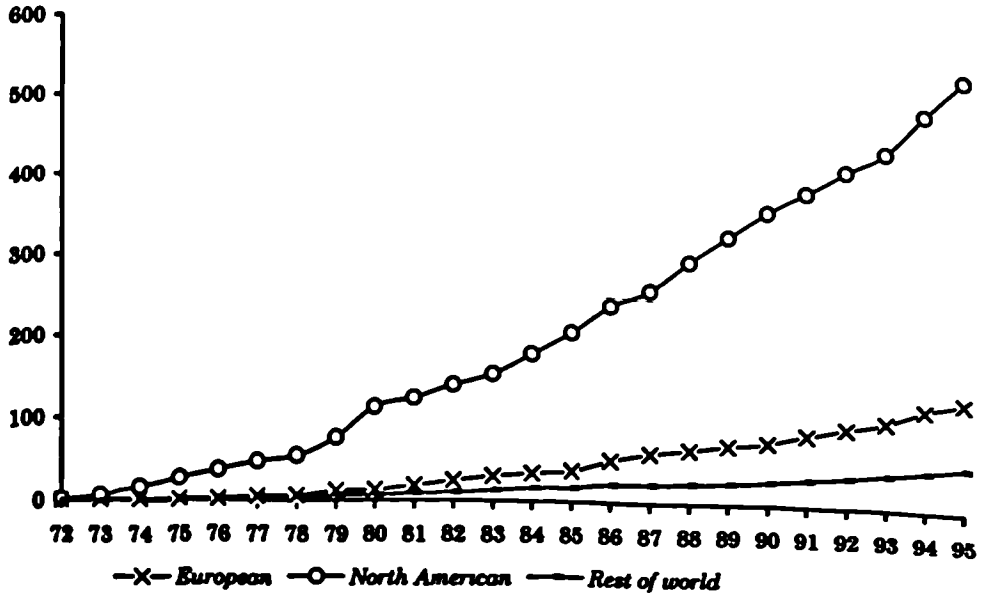
Figure 6.3: Distribution of articles by geographical provenance



The European Union category is largely accounted for by articles from British (10.9 per cent), German (3.0 per cent) and European Commission (1.8 per cent) authors. The remaining category includes articles from Australasia, South East Asia, Japan, the Middle East, India, Latin America, and those countries of Europe currently outside of the Union.

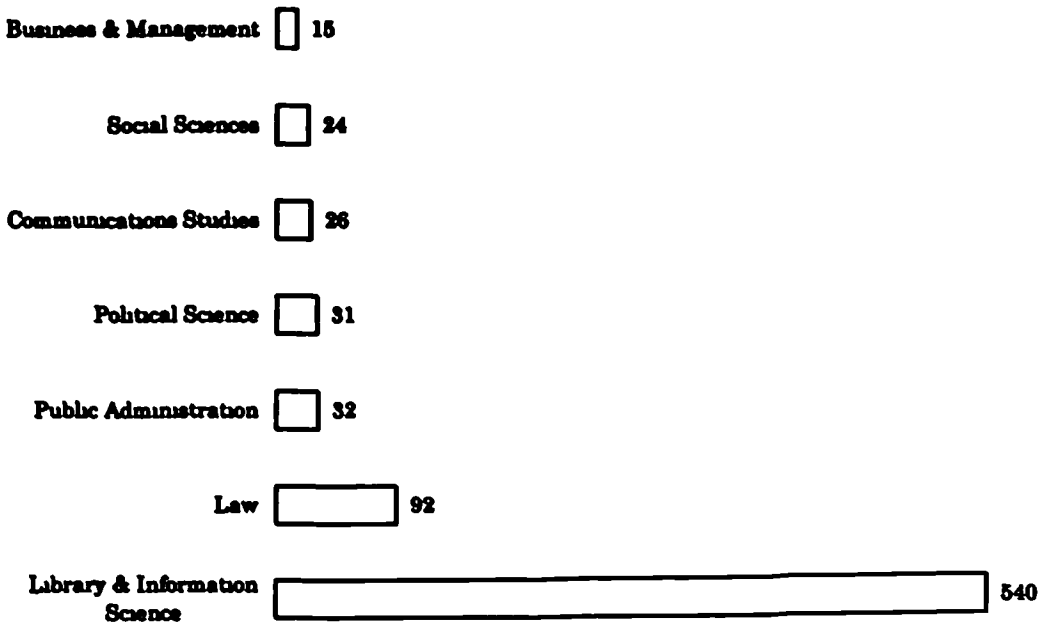
The attention devoted to information policy by North American scholars is further reinforced by Figure 6.4 overleaf which shows that cumulative article production has been consistently growing at a faster rate than in Europe or elsewhere since 1972.

Figure 6.4: Cumulative growth of articles by author region



Where information policy authors publish: the distribution by journal category

Figure 6.5: Distribution of articles by journal category



The disciplinary profile of the test collection is represented in Figure 6.5 as a distribution of articles by *ISI* journal category. The *ISI* journal category is assigned *at the level of the journal title* and so it does not provide any indication of the subject nature of individual articles. Figure 6.5 should therefore be

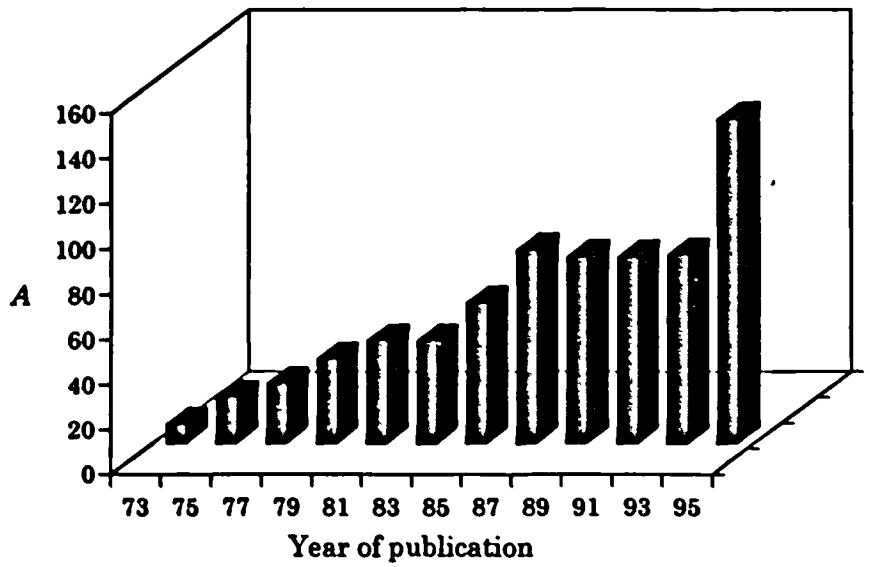
interpreted as showing the kinds of serials in which information policy articles are published, and no construction about topics or author affiliations should be inferred.

The main destination for information policy articles is the library & information science category, with further representation across a broad canvas of legal, political and social science titles. This empirical evidence offers some limited support for the often-rehearsed view that information policy is multidisciplinary in scope and character.

6.3.2 Patterns of growth in the test collection

The distribution of article production over time shows that there has been a sustained and growing interest in information policy topics over the period 1972 to 1995:

Figure 6.6: Article production in two-year increments



The information policy serials literature appears to be doubling in volume approximately every six years (Figure 6.7 overleaf). Growth patterns of this nature are often observed in bibliometric studies.

Figure 6.7: Cumulative article production[24]

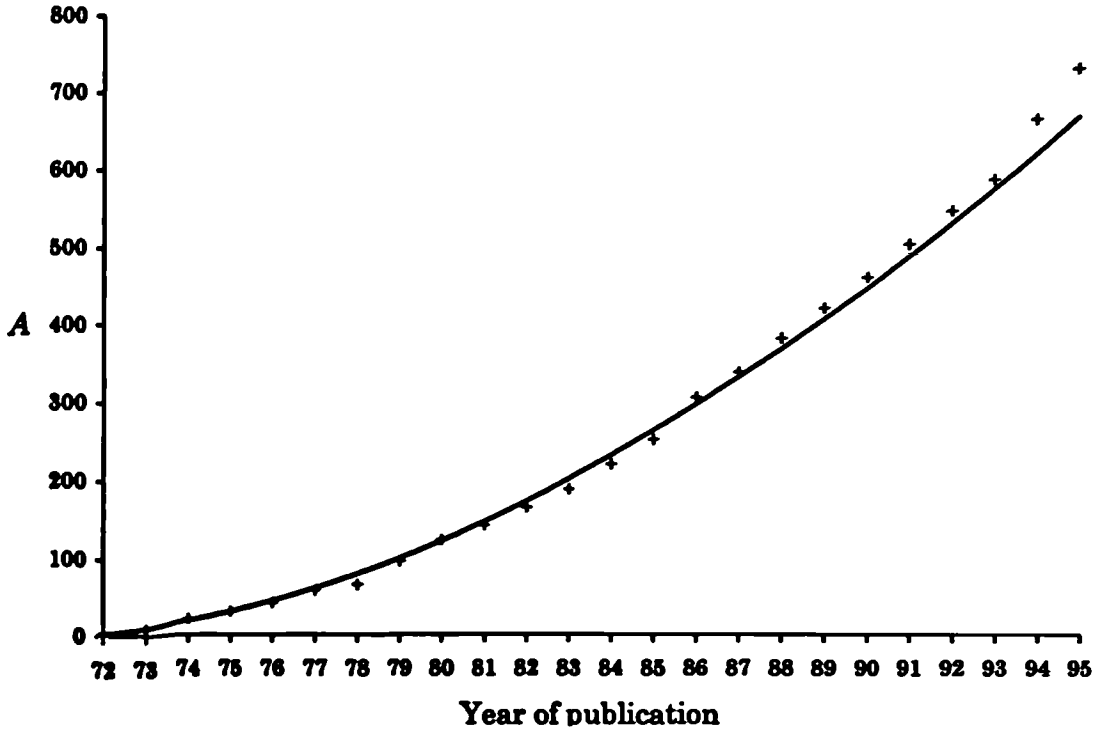
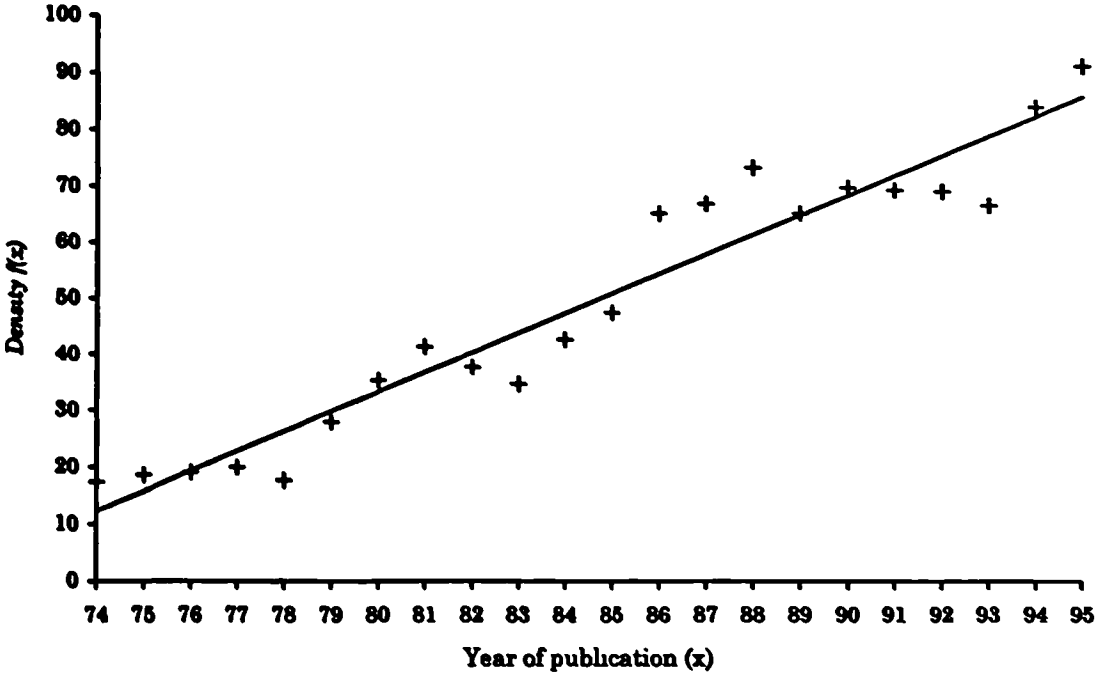


Figure 6.8: Relative article production



[24] Power model ($R^2 = 0.998$). $A = 2.693x^{1.735}$

Given the question mark noted earlier over the possible instability of *ISI* editorial policy over time, cumulative article production is represented in Figure 6.8 as a proportion of all *Social SciSearch* records. The data are expressed as bibliography entries per 100,000 *Social SciSearch* records, and presented as a three-year moving average. The trendline[25] suggests that the volume of information policy articles is expanding in both absolute and relative terms. In other words, information policy is capturing greater 'market share' within the total population of records indexed by *Social SciSearch.*

6.3.3 Knowledge accumulation

These observations on the growth of the information policy serials literature are interesting, but they tell us nothing about the *processes* underlying that growth. This section considers two aspects of knowledge accumulation in the information policy serials literature: the 'scholarliness' of the contributions laid down, and the age distribution of the supporting (cited) literature.

The 'scholarliness' of the test collection

The test collection is broadly-based, encompassing research-based and opinion articles and articles by academics and practitioners. Windsor & Windsor (1973) argue that the ratio of source papers without references to those with references offers a measure of the maturity and 'scholarliness' of a given literature. In the case of the information science literature, Windsor & Windsor found this ratio to be 30:70[26].

The information policy test collection performs 'very well' by this standard, with a ratio of 13:87 across the test collection (Figure 6.9 overleaf), although this possibly says as much about *ISI* editorial standards as offering any fundamental insights into the structure of the information policy literature.

---

[25] Linear regression model ($R^2 = 0.927$). $y = 3.484x + 8.874$

[26] Kajberg (1996) examined four Danish library journals over the period 1957-86 and returned a value for the 'Windsor ratio' of 20:80.

Figure 6.9: The 'Windsor ratio'

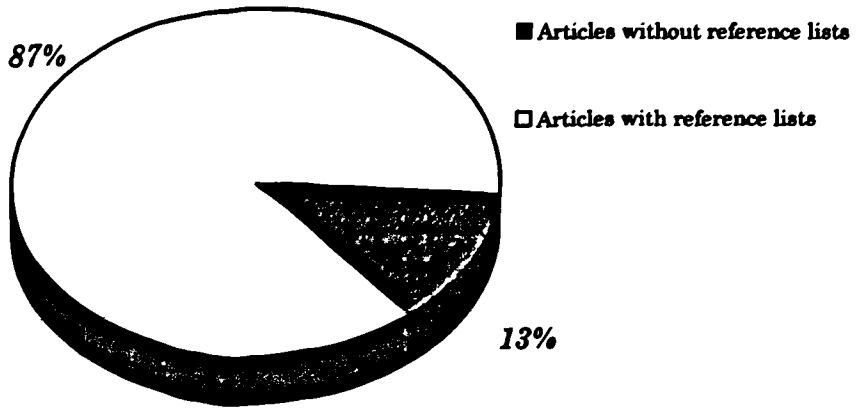
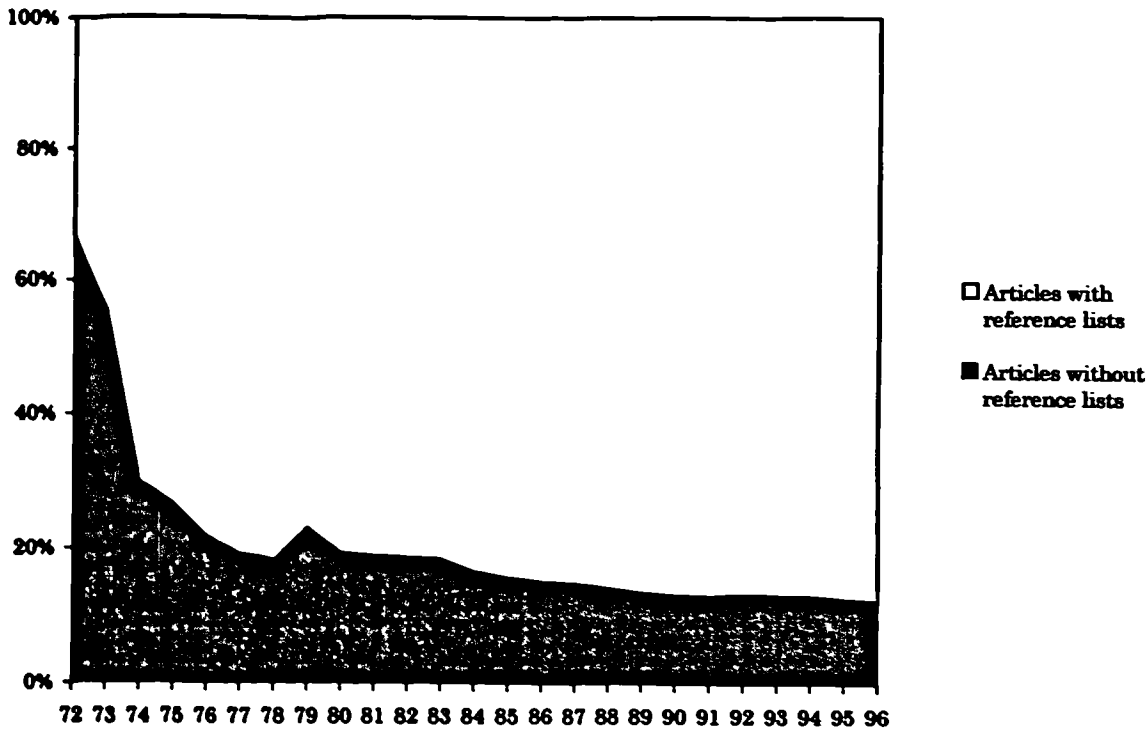
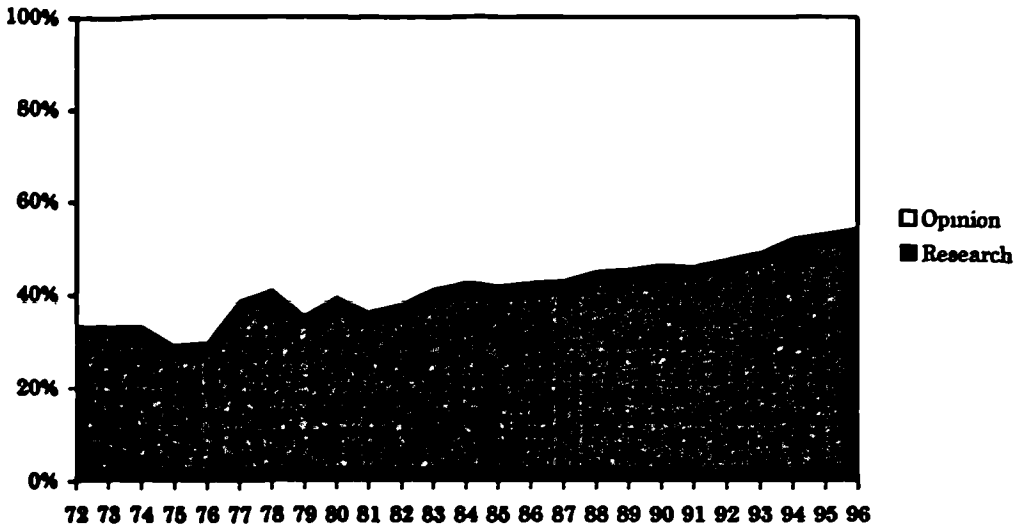


Figure 6.10: Changes in the 'Windsor ratio' over time



The data in Figure 6.10 suggest that the articles in the test collection are showing an increasing propensity to include citations. Kajberg observed a similar phenomenon over a 29-year period in relation to four Danish library journals, but conceded that "obvious explanations for these movements do not present themselves" (Kajberg, 1996:77). One factor which may be relevant in the case of the information policy literature is a shift in the balance of opinion and research-based articles (Figure 6.11 on the next page).

Figure 6.11: Comparative growth rates: research *vs* opinion articles



The age distribution of the supporting literature

The Windsor ratio reports on just one aspect of how authors interact with existing knowledge: through explicit citation. It tells us nothing about the nature of their use of the accumulating body of recorded literature. Information scientists have long been interested in the differences in the way that knowledge accumulates in various subject areas. One important aspect of knowledge accumulation is the extent to which older materials are knitted into the fabric of more recent publications through citation. As long ago as 1953, Stevens claimed that science and technology:

> *"exhibit a high concentration of papers in a select nucleus of special journals, and also in a brief span of time covering a few current years. In contrast, the literatures of the social sciences and humanities exhibit a great dispersion of publications in different forms, on different subjects, and over a comparatively long span of time"* (Stevens, 1953:12).

Price also noted that science distinguishes itself from other fields of study in the way that scientists refer to their literature. If references to the existing literature were distributed evenly across the entire archive of material which was available to be cited, he reasoned that the age distribution of citations in any one year should reflect the age and size of the archive. Instead, he found

that more recent papers in science were much more likely to be cited than his simple probability model suggested.

In a 1970 article, he introduced a new bibliometric indicator, Price's index (PI). This is given by:

$$PI = (n_1/n_2)*100$$

where n_1 is the number of cited references with a relative age of less than six years[27] and n_2 is the total number of cited references.

Price found that the value of the index varied according to the discipline or field of study under investigation:

> *"Perhaps the most important finding I have to offer is that the hierarchy of Price's index seems to correspond with what we intuit as hard science, soft science, and nonscience as we descend the scale. Biochemistry and physics are at the top, with indexes of 60 to 70 percent, the social sciences cluster around 42 percent, and the humanities fall in the range 10 to 30 percent"* (Price, 1970:4)

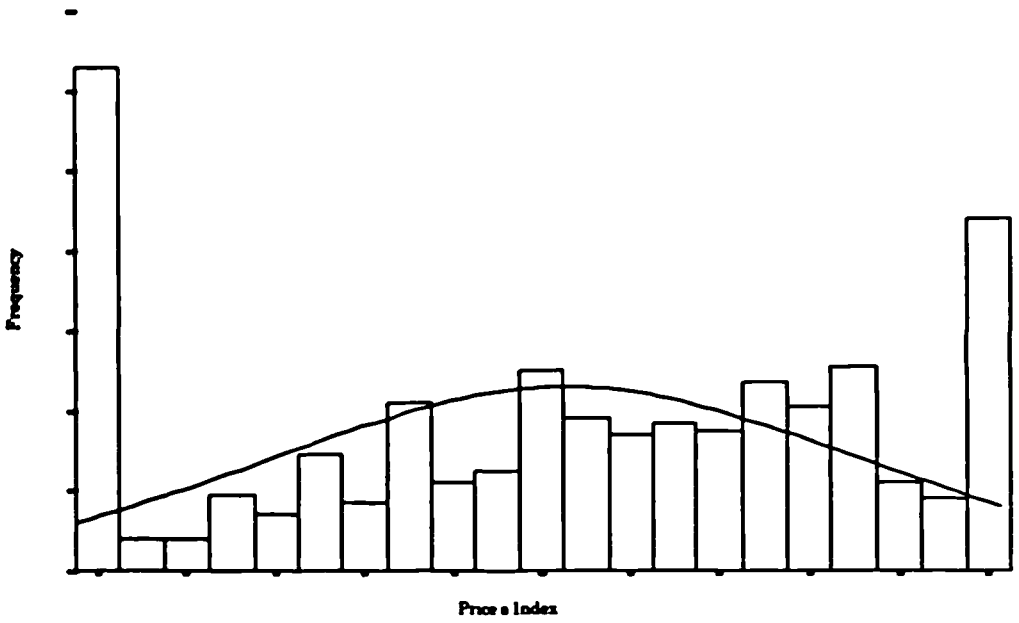
Price dubbed this phenomenon 'The Immediacy Effect' and further proposed the existence of two main types of literature: the 'ephemeral' and the 'classical'.

The value of Price's index may be determined in two ways. Price himself used a global measure while Moed (1989) proposed an improvement by calculating the average index value across a population of articles. Figure 6.12 shows the frequency distribution of Price's index across the information policy test collection. The distribution approximates to a normal distribution overlapping with two sub-populations: one with an index of 0 per cent and another with an

[27] Egghe & Rousseau (1995) observe an ambiguity in relation to how Price's index is calculated in the published literature. Some authors, notably Price, state that they use 'the first five years'. In this terminology it is unclear whether the year of publication, d, is year zero or year one. Moreover, it is unclear whether or not this year is included. In this thesis, Price's index is calculated conservatively on the basis that $d = 1$.

index of 100 per cent. Moed excluded these two sub-populations in calculating his 'corrected' value for the index.

Figure 6.12: Frequency distribution of Price's index

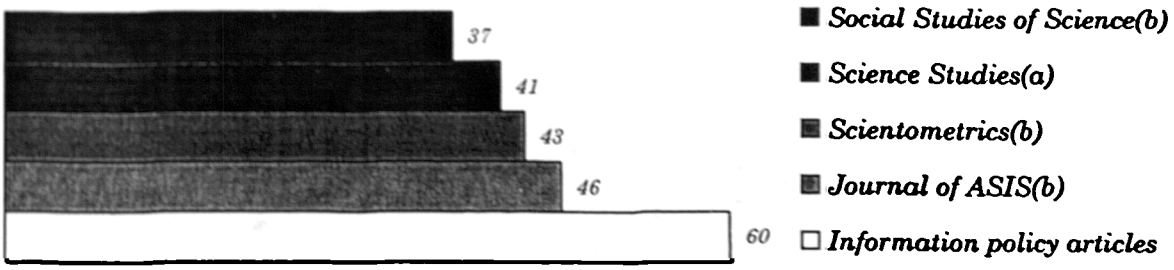


The values for Price's index for the information policy test collection are thus:

Price's global measure	60 per cent
Moed's 'corrected' average	57 per cent

Whichever way the index is calculated, the value is surprisingly high, as the comparative data in Figure 6.13 below testify:

Figure 6.13: Comparison of Price's index with other studies

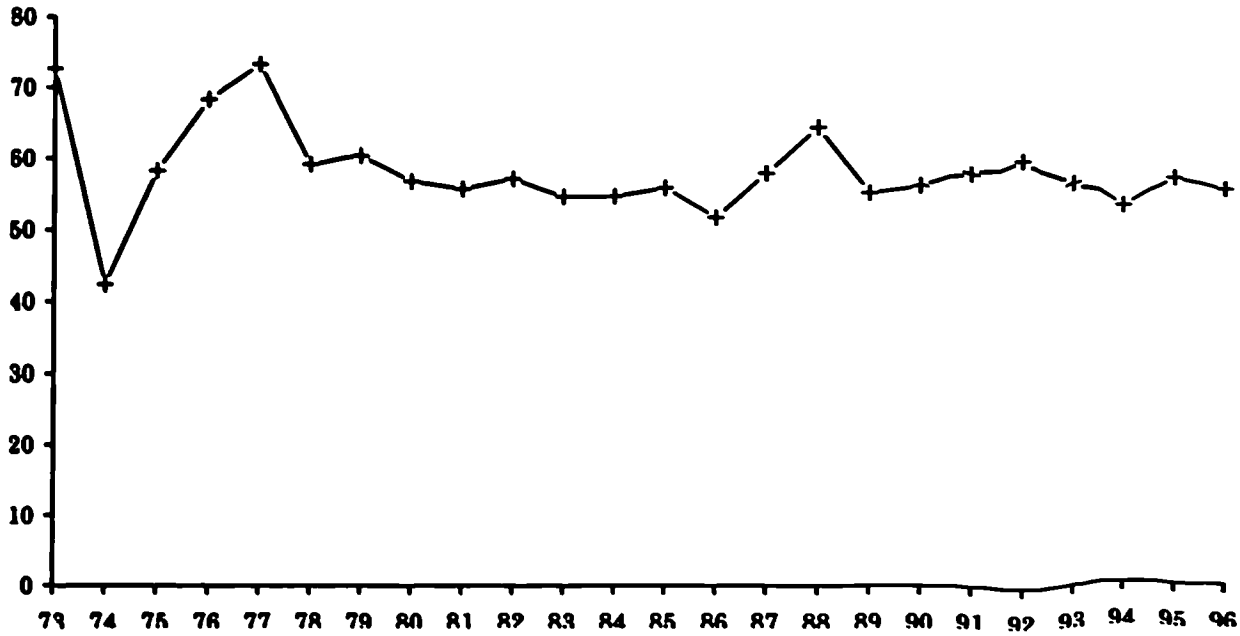


(a) Spiegel-Rosing (1977)
(b) Wouters & Leydesdorff (1994)

The data presented in Figure 6.13 compare Price's index for the information policy test collection with available published figures for four social science journals. The value for Price's index is much higher than one might have anticipated; indeed it suggests that the citation practices of information policy authors more closely resemble those of workers in the natural rather than the social sciences.

One possible explanation for this apparent anomaly is that information policy is a relatively small field (smaller than scientometrics?) with a consequently small archive to draw upon. However, Price's index shows a high degree of stability over time, as Figure 6.14 illustrates, and appears to be a structural feature of the literature:

Figure 6.14: Stability of Price's index over time



What conclusions can be drawn from these results? Information policy is evidently not a hard science, although its authors do seem to share with, say, biochemists, a tendency to emphasise more recent works in their reference lists. In developing an explanation of his empirical data, Price reached for the major concept available at the time (1970) to describe differences between disciplines: the hierarchy of the sciences introduced nearly two centuries earlier by Comte. This hierarchy emphasises cognitive structure and it perhaps sits uneasily with

the essentially sociological patterns of behaviour which Price was observing. Particularly incongruent here is Price's assertion that that it is the concerted attention of scientists that produces the immediacy effect:

> *"... if you want to make the field firm and tight and hard and crystalline you have to play with your peers and keep on the ball by citing their recent work"* (Price, 1970:15).

In other words, a relatively high degree of consensus, more usually the case than not in the hard sciences, might be expected to be a good predictor of 'high' PI values. The tacit assumption of Price's work and those that have followed is that such high levels of consensus do not obtain in the social sciences or arts and humanities. In a major review article, however, Cole (1983) rejects this notion and argues that in some cases, social sciences do exhibit high levels of consensus. Cole further suggests that the fundamental differences between disciplines lie not in citation habits but in the structure of their knowledge systems, particularly in relation to how empirical knowledge is codified into "succinct and interdependent theoretical statements" (Cole, 1983:112).

Intriguingly, Cozzens (1985) points out that several cocitation studies of the development of disciplines have found a relationship between immediacy and periods of intellectual focus[28]:

> *"At times of intellectual excitement, whether in response to a recent breakthrough which has opened up a new set of interesting questions or because researchers are trying to resolve a theoretical issue, members of the speciality concentrate their references more heavily on a few papers, and the literature they cite tends to be more recent than it is at other times"* (Cozzens, 1985:436).

---

[28] For example, Small (1977) found that immediacy rose during a 'mini-revolution' in collagen research. Sullivan, White & Barboni (1977) demonstrated a similar phenomenon during a period of rapid theory change in the physics of weak interactions. Dean (1980) found that psychologists generally cited papers about seven years old: however, this dropped sharply during the period when additivity theory was displacing inhibitory theory as the dominant explanation for the probability of a conditioned response. Price's index declined to its normal rate only after the after the controversy had been resolved.

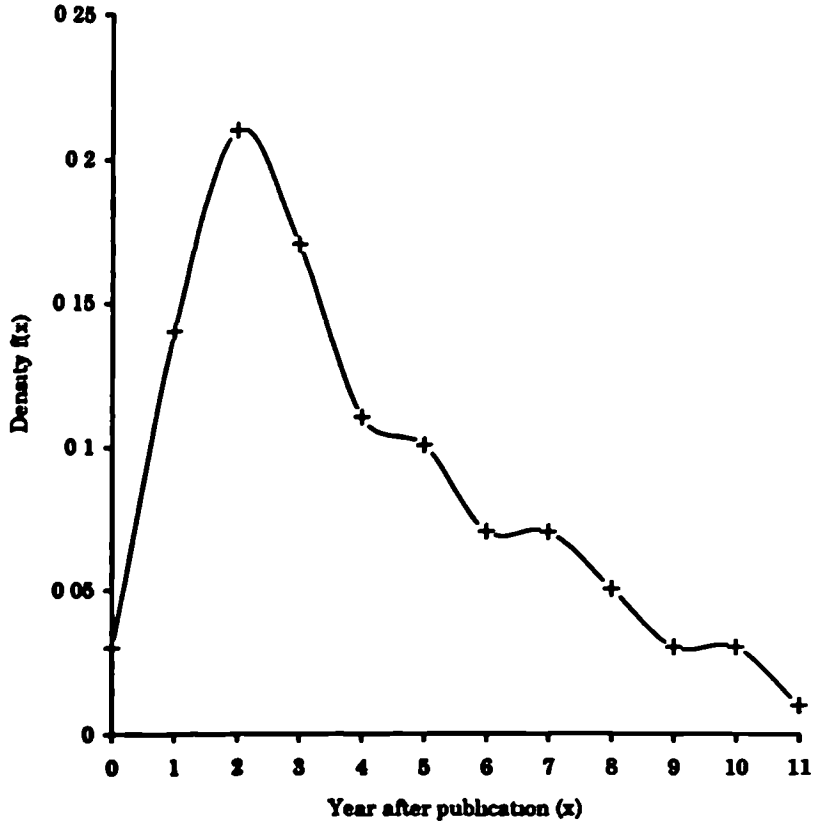
In a 1979 study, Small & Crane compared the literature of high energy physics with that of three social science disciplines: psychology, economics and sociology. The authors' argument was that if the processes of knowledge growth in the social sciences are similar to those in the natural sciences, then it should be possible in both cases to find clusters of cocited documents representing groups of researchers pursuing closely related problems. Small & Crane showed that the literatures showed differing propensities to cluster, with the highest propensity exhibited by the most 'immediate' literatures (high energy physics and psychology).

The results of these studies suggest that a causal explanation for Price's immediacy effect is likely to be sociological in its origin. The large PI values associated with the information policy literature may be indicative of a higher degree of consensus and intellectual focus than anticipated, rather than pointing to a cognitive structure which is similar to the higher levels of the Comte's hierarchy of the sciences. Some of these themes are pursued further in the author cocitation analysis in Chapter 8.

6.3.4 Ageing and obsolescence of information policy articles

The material in the previous section related to the use made of the supporting literature by authors represented in the information policy test collection. In this section, the attention shifts to how articles in the bibliography have been cited by others. The main objective is to examine the ageing of the information policy serials literature by seeing how citations are distributed in time. While citations alone cannot to depict the totality of the reception and ageing processes (data on document use would be needed to explore this fully), the distribution at least indicates how information policy articles are received by one group at one point in time: other authors preparing their own work for publication. The data in Figure 6.15 overleaf were compiled by searching for the citations to date (299) to the 127 bibliography entries published over the period 1981-85.

Figure 6.15: Distribution of cited information policy articles over time



One characteristic of a rapidly ageing as opposed to a more classical literature is a skewed distribution and a relatively small median value. The pattern displayed in Figure 6.15 suggests that information policy articles enjoy a period of rapid citation over the first three years (reception), followed by a relatively slow decline (obsolescence) over the following eight or nine years[29]. The citation half-life for the test collection articles is three years; unsurprising in the light of the high Price's index for the literature, but certainly more typical of the patterns usually observed in the natural sciences. Bottle & Gong, for example, returned an identical estimate (three years) for the citation half-life of the biochemistry literature (Bottle & Gong, 1987).

[29] In the typology of Glänzel & Schopflin (1995), this is characteristic of a 'Type III' ageing process, and may be indicative of a bibliography which is heterogeneous in respect of its reception and obsolesence characteristics.

6.3.5 Bradford studies

Attempts to identify 'core zones' in bibliographies of serials articles can be traced to the work of Bradford (1934). In this section, the main corpus is partitioned into zones in order to identify the most highly productive serial titles.

Concepts of core and scatter

The Bradford concepts of 'core' and 'scatter' are important for this thesis because they are to some extent indicative of the underlying social and intellectual structure of the field of study. Chubin observes that:

> *"If there was no 'scatter', scientists would be divided into small groups sharing the same interests, speaking only to each other, and reading and citing only each other's work ... Both [core and scatter] are necessary, the former to permit scientific knowledge to cumulate and grow, the latter to prevent it from becoming a ... sect-like phenomenon"* (Chubin, 1976:472).

In less highly structured or specialised disciplines there is a general expectation that people will read widely outside of their own current areas of concern. In arts and humanities subjects, for example, potentially relevant ideas may be gleaned from a very wide variety of sources. The breadth of influences to which a researcher is receptive is of course a function both of personal inclination and of disciplinary conventions. Nadel suggests that catholicity of interests is also a function of the maturity of a specialism. He noted that in its early stages, research on superconductivity was characterised by a dispersion of articles in a variety of journals but that it later became 'institutionalised' by the increasing concentration of published material in a relatively limited number of specialist sources (Nadel, 1980). A tentative causal explanation for this effect is that it becomes more and more difficult for the typical researcher to stay abreast of developments across an increasingly specialised field because of the inaccessibility of language and logical structures in the adjoining literatures.

Bradford (1934) was the first to draw attention to statistical regularities in the distribution of articles across journal titles in (reasonably comprehensive) subject bibliographies. His method is replicated here with the assistance of *The Bibliometric Toolbox* software. Most of the published empirical investigations into Bradford's Law have come from the natural and medical sciences and only rarely from the social sciences (see, however, Coleman, 1992).

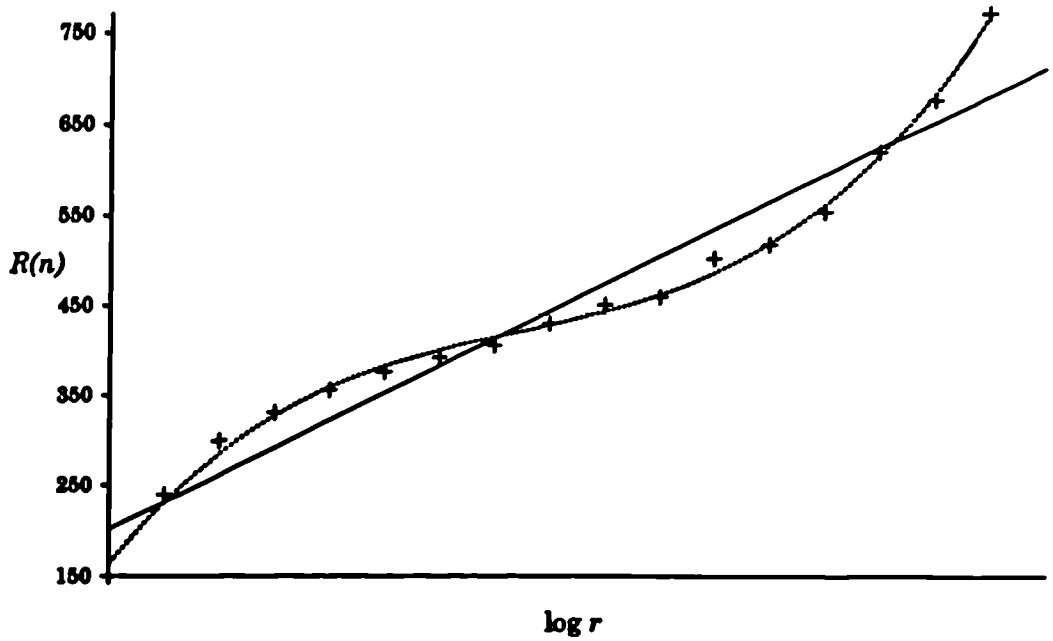
Table 6.2 illustrates the presence of a relatively small number of highly productive core titles and a long tail of journal titles which contribute few articles, in many cases only one.

Table 6.2: Bradford ranking of *M* corpus journals

J	A	Total articles	Cumulative titles, *r*	Cumulative articles, *R(n)*
1	150	150	1	150
1	90	90	2	240
1	60	60	3	300
1	30	30	4	330
1	24	24	5	354
1	20	20	6	374
1	16	16	7	390
1	13	13	8	403
2	12	24	10	427
2	10	20	12	447
1	8	8	13	455
7	6	42	20	497
3	5	15	23	512
9	4	36	32	548
22	3	66	54	614
30	2	60	84	674
97	1	97	181	771
181		*771*		

The data in Table 6.2 are represented graphically on the next page as a 'bibliograph' (Figure 6.16). This plots the logarithm of a given journal's rank, *r*, against the cumulative number of articles up to that point, *R(n)*.

Figure 6.16: Bibliograph of *M* corpus journal productivity



The solid trendline is a linear regression model[30] while the dotted trendline represents a curvilinear expression[31].

Much attention in the bibliometric literature has focused on the generalisability of Bradford's Law and, in particular, on the goodness of fit with linearity obtained in experimental bibliographs. Departures from linearity have attracted some attention and attempts at explanation (see, for example, Drott, 1981). In a 1993 paper, Coleman suggested that there is a relationship between linearity and the homogeneity and completeness of a bibliography. It will be noted that the curve obtained here is definitely *S*-shaped as opposed to the more usual 'classical' *J*-shaped distribution of many published bibliographs.

One might expect *a priori* that 'information policy' might best regarded as a heterogeneous rather than a homogenous construct. Berelson (1960) draws a useful distinction between *data-specialties* and *word-specialties*. Data specialties (experimental psychology, for instance) are those which are characterised by distinctive, public procedures, standard methodologies and special apparatus. Word specialties (such as information policy) are, in contrast, less tightly defined

[30] **Linear regression model** ($R^2 = 0.937$). $R(n) = 29.79 \log r + 172.23$
[31] **Third order polynomial model** ($R^2 = 0.997$). $R(n) = 0.381(\log r)^3 - 9.673(\log r)^2 + 77.412$

in terms of topic, procedure, institutional structures, purpose or even in terms of a consensus around the paradigms employed.

Coleman argues that the bibliometric characteristics of word specialties differ from data specialties, since the latter enjoy:

> *"a level of standardisation that reduces ambiguity in the task of making distinctions between publications that are proper to that speciality and those that belong to neighbouring specialties ... A word-based speciality does not enjoy the same capability because, although it may have a somewhat distinctive vocabulary, concepts are more easily borrowed by and from neighbouring word-based specialties than are apparatus and procedures. Such movement of concepts contributes to a blurring of boundaries between word-based specialties and makes it necessary for both the bibliographer and the journal editor to distinguish properly among the publications of cognate specialties. As a result ... the journal editor finds the topical hierarchy of the journal fuzzy and hard to obey. Thus, the bibliography of a word-specialty ought to show less concentration than that of a data-specialty"* (Coleman, 1993:88).

Information policy certainly appears to sit well with Berelson's conception of a word-specialty; the term is an umbrella designation embracing a variety of topics and approaches (the elucidation of which is a central aim of this thesis). Coleman's remarks above also point to the difficulties of obtaining a complete bibliography for a word-based specialty. As noted in Chapter 5, attempts to maximise the recall obtained in the experimental bibliography were made pragmatically: no formal indication of recall has been attempted. *The Bibliometric Toolbox* offers a tool for estimating the size of the total literature, based on the work of Egghe (1989). This procedure requires the determination of the point on the bibliograph where the Groos droop (the point where the relationship between journals and articles begins to diminish). Unfortunately, a Groos droop is not evident in the bibliograph (Figure 6.16) and so no attempt has been made to estimate the theoretical size of the information policy literature.

Partitioning the test collection

Having generated a Bradford ranking, the next stage of the analysis was to effect a partition of the journals into zones of productivity (information which is carried forward into the cluster analysis reported in the next Chapter). This is well understood to be an imprecise procedure and so four approaches were made in an attempt to find the optimal partitioning of the data:

1. The first approach was to inspect the Bradford ranking visually and to divide the journals into zones of roughly equal productivity, such that they contained the smallest equal number of articles necessary to effect a Bradford partition.

2. Goffman & Warren (1969) developed a method for determining the size of a minimum Bradford cohort. This is just larger than half the number of singleton journals (i.e. the journals that produce just one article each). In this case, there are 97 singletons, so the Goffman/Warren minimum Bradford zone is (97 / 2) + 1 or 49.

3. Egghe (1989) introduced a calculation for the maximum number of Bradford zones p in a partition:

$$p = 2 \ln (1.781 Y_m)$$

where Y_m is the number of articles for the most productive journal.

In this case $p = 2$ in (267.15) = 11.18, which approximates to 11.

4. Finally, Brooks (1990) developed the Minimum Perfect Bradford Partition (MPBP) method and implemented this in the design of *The Bibliometric Toolbox*. This method uses the power of the computer to exhaustively examine the Bradford ranking for cohorts of articles that satisfy the twin hallmarks of a Bradford partition: an equal number of articles per zone, and an increasing number of journals.

The partitioning which resulted from using the first three manual operations above is shown in Table 6.3:

Table 6.3: Manual partitioning of journals in *M*

Bradford zone	*J*	*A*	Bradford multiplier
1	1	150	
2	2	150	2.00
3	9	147	4.50
4	42	167	4.67
5	127	157	3.02
All zones	*181*	*771*	*3.55*

The ratio between the number of periodicals in the nucleus and the number in each succeeding zone was then estimated; this is the Bradford multiplier b_m for journals in the main corpus.

These results may be compared with the results of partitioning the data by Brooks' MPBP method using *The Bibliometric Toolbox* in Table 6.4. This method yielded six possible cohort sizes that would effect a reasonable Bradford partition of the data. These were (with remainders in parentheses):

374 (23) 354 (63) 330 (111) 329 (111) 240 (51) 150 (21)

Table 6.4: Computer-assisted partitioning of journals in *M*

Bradford zone	*J*	*A*	Bradford multiplier
1	1	150	
2	2	150	2.00
3	10	150	5.00
4	38	150	3.80
5	114	150	3.00
remainders	16	16	
All zones	*181*	*771*	*3.45*

This analysis confirmed 150 as the optimal cohort size. The practical utility of Brooks' MPBP method is very limited, however, since the software allocates journals of equal productivity into different zones.

Finally, *The Bibliometric Toolbox* computes a number of useful indicators (Figure 6.15) which can be used to assess various characteristics of the partitioning and of the literature in general[32].

Figure 6.17: Key statistics associated with Brooks' MPBP partition

Cohort size	*150*
Goffman/Warren minimum cohort	*49*
Clustering index	*13.76*
Average recurring productivity (ARP)	*8.02*
Egghe's C prime index	*0.003*
Goffman/Warren mean multiplier	*3.45*

The Bradford core journals

The minimal nucleus for the test collection consists of a single journal, *Government Information Quarterly.*

---

[32] **Clustering index.** The clustering index (CI) expresses how intensely a given literature is clustered. It is a constant, and is the result of dividing the total production of the recurring (i.e. non-singleton) journals by the Goffman/Warren minimum cohort. Brooks (1990) has observed that strongly clustered literatures yield CI values greater than unity, while weakly clustered literatures have CI values of less than unity. The value obtained here, 13.76, suggests that the experimental bibliography is strongly clustered.

Average recurring productivity. The average recurring productivity (ARP) expresses the average production of the recurring journals of a bibliography. It is found by dividing the total number of articles in recurring journal titles by the number of recurring journal titles. The ARP complements the clustering index since it is possible for literatures to return similar CI values but to cluster in different patterns: one literature may be composed of a few very productive journals while another is composed of many journals recurring only two or three times. In this example, the ARP of the first literature would be much greater than the second. Taken together the clustering index and the average recurring productivity express the degree of intensity of the clustering of the literature along two dimensions: total contribution of the recurring journals to the singletons and the average production of the recurring journals.

Egghe's C prime index. This index expresses the extent to which a literature is evenly partitioned over the zones created. Its value lies in the range 0.0 to 1.0 and if every zone has an equal number of articles, then C prime is zero.

Goffman/Warren mean multiplier. This indicator provides an estimate of the Bradford multiplier b_m. This describes the relationship between the number of journal titles producing the nucleus and the number in each succeeding zone. This constant varies between literatures and depends on the number of zones into which a literature is partitioned. As the number of zones increases so the Goffman/Warren mean multiplier decreases, although it always remains greater than unity.

The two journals in Bradford zone 2 are (in decreasing order of productivity):

Journal of Government Information
(incorporating *Government Publications Review*)
Telecommunications Policy

The journals in Bradford zone 3 are (in decreasing order of productivity):

Aslib Proceedings
Proceedings of the Annual Meeting of ASIS
Journal of Information Science
Public Administration Review
International Forum on Information and Documentation
Duke Law Journal
Information Age
Administrative Law Review
Journal of the American Society for Information Science

The twelve most productive journals in the bibliography thus comprise only 6.6 per cent of all titles but yield 58.0 per cent of all articles. Eight of the twelve most productive journals occupy a central position in the library and information science literature. The other four (*Telecommunications Policy*, *Public Administration Review*, *Duke Law Journal* and *Administrative Law Review*) could hardly be categorised in this way although the potential relevance of these titles to the field of information policy is evident.

Table 6.5 on the next page shows a breakdown of the test collection journals by Bradford zone and *ISI* journal category.

Table 6.5: *M* corpus journals by Bradford zone and *ISI* journal category

ISI journal category	zone 1	zone 2	zone 3	zone 4	zone 5
Information & library science	1	1	6	24	19
Communications studies		1		3	5
Law			2	7	41
Public administration			1	3	4
Politics / international relations				1	17
Other				4	41
All categories	*1*	*2*	*9*	*42*	*127*

This analysis suggests that the core information policy journals are mainly library and information science titles and that other journal categories contribute disproportionately to the outer Bradford zones. This is an interesting finding in relation to the debate about whether information science is indeed the most natural home discipline for information policy studies (see, for example, Burger, 1993).

6.4 Patterns of authorship

6.4.1 Author productivity

The records in *Social Science Citation Index* are indexed by first-named author only. The discussion which follows is based entirely on patterns observable in the distribution of these authors; while this means that the results are distorted, the decision to proceed on this basis was based on practical considerations. For instance, it facilitates comparison with the many existing bibliometric studies which share the same limitation. It should also be noted that the incidence of collaborative authorship in the test collection is very low compared to other literatures, and a visual inspection of a listing of second and subsequent named authors suggested that the vast majority of these individuals would only be represented by a single article.

A listing of the top ranking authors represented in the bibliography is shown in Table 6.6. The ranking is dominated by American authors: only one of the top ten most productive authors is based outside the USA: Robin Mansell (of the

Science Policy Research Unit, University of Sussex). Notable also is the relatively recent entry of several of the authors.

Table 6.6: Top ranking personal authors by article production

Rank	Name of author	First article	Cum. % all articles
1	Harold C. Relyea	1977	2.33
2	Peter Hernon	1986	3.89
3	Timothy J. Sprehe	1984	5.19
4	Charles R. McClure	1987	6.10
5	Sandra Braman	1989	6.74
6 =	Joan C. Durrance	1982	7.26
6 =	J.C. Griffith	1989	7.78
6 =	Robin Mansell	1985	8.30
6 =	H B. Shill	1989	8.82

The purpose of this ranking exercise was to generate indicators for inclusion in the content analysis phase: essentially so that questions of the form, "Are there differences between highly productive and less highly productive authors?" could be asked.

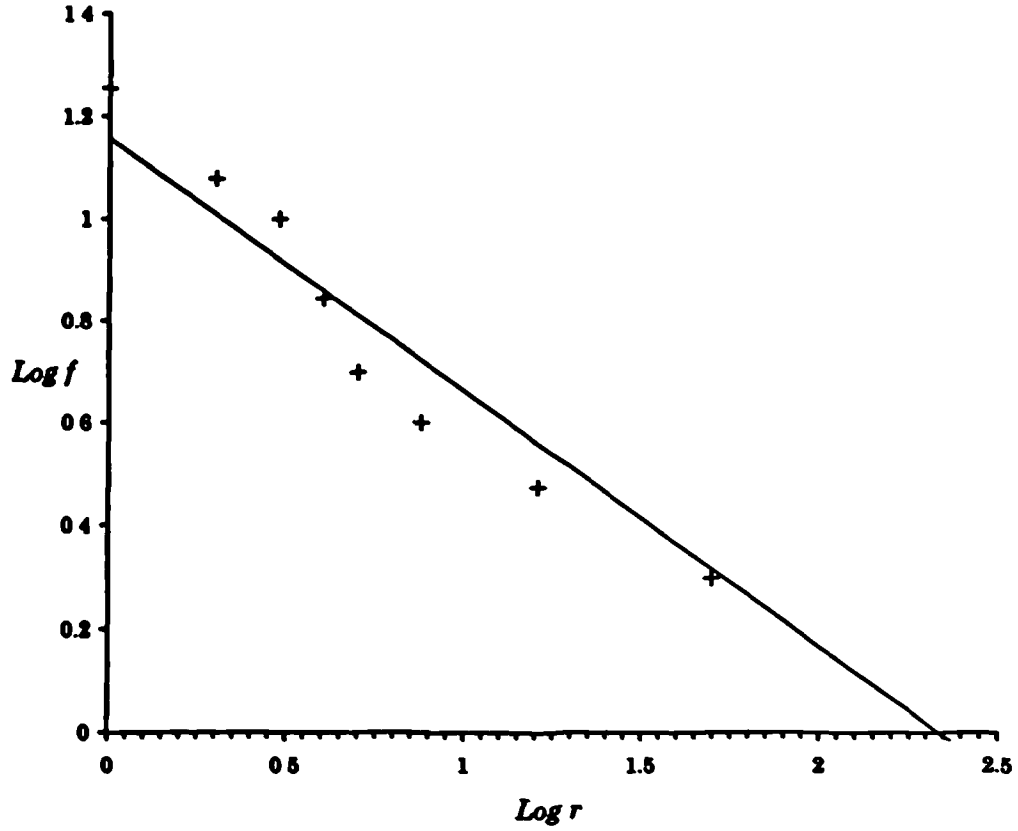
The distribution of author productivity is shown in Table 6.7 overleaf. The most notable feature is the striking proportion of authors who are represented by a single article (72 per cent). The final column, rank, is included to facilitate the graphical representation of the data as a Zipfian rank-frequency distribution in Figure 6.18.

One of the features of *The Bibliometric Toolbox* is that it enables an analysis of author productivity to be accomplished very easily. In 1926, Lotka proposed a distribution that modelled the frequency distribution of scientific productivity ('Lotka's Law'). Lotka's law takes the form $y_x = c/x^n$, where y_x is the number of authors credited with x (1,2,3 ...) papers in a given literature, c is the number of authors writing *one* paper and n is a rate. It is possible from Figure 6.18 to estimate a value for the exponent, n (in this case, $n = 2.34$).

Table 6.7: Distribution of author productivity

Frequency, *f*	Cohort size	Articles	Rank, *r*
18	1	18	1.0
12	1	12	2.0
10	1	10	3.0
7	1	7	4.0
5	1	5	5.0
4	4	16	7.5
3	13	39	16.0
2	54	108	49.5
1	556	556	354.5
	632	*771*	

Figure 6.18: Author productivity: Zipfian rank-frequency chart[33]



Lotka's Law is sometimes called the 'inverse square law of scientific productivity', based on the assumption that $n = 2$. Lotka's Law predicts that the number of authors making two contributions is about one-quarter of those making one; the number making three contributions is about a ninth, and so on.

The empirical straight line relationship found suggests that the fit with Lotka's Law is good. The exponent n provides an indicator of the concentration of

[33] Linear regression model ($R^2 = 0.95$). $y = -0.495x + 1.158$

author contributions to the bibliography: lower values of n are associated with concentrations of authors who have multiple papers. Average productivity is thus greater since more papers are distributed over relatively fewer authors.

Nicholls (1988) conducted a series of experiments on 100 datasets in an empirical investigation of Lotka's Law. He found that the values of n increased significantly when second and subsequent authors were included. The comparative data in Table 6.8 are drawn from the work of Nicholls and relate to his findings for first-named authors.

Table 6.8: Comparison of Lotka studies

Literature	Mean n	Source
Natural sciences	2.20	Nicholls (1988)
Information policy	2.34	Rowlands
Social sciences	2.49	Nicholls (1988)
Humanities	2.55	Nicholls (1988)

(adapted from Nicholls, 1988).

This analysis suggests that information policy sits somewhere between the natural and social sciences in terms of author productivity.

6.4.2 Collaboration with other authors

Patterns of collaborative authorship can offer further insights into the cognitive, social and institutional organisation of a discipline. In some fields, publications are likely to bear the names of two or more authors; in others, sole authorship is the norm. On the face of it, an individual academic's reputation is likely to be most decisively established if the person concerned takes full, unambiguous responsibility for his or her own work. There are however some very good reasons for co-authoring publications. In some areas of 'Big Science', the scale and complexity of the experiments and the limited availability of apparatus is such that team-working is the only realistic strategy. Similarly, certain highly multidisciplinary problems in the social sciences may necessitate a division of labour. A study of collaborative patterns of working in four areas of the social sciences (sociology, psychology, economics, political science) by Fox & Faver concluded that division of labour is the most common strategy for:

> *"routine research activities, those under time pressure, and those demanding complementary competencies ... [but] more complex tasks may benefit from some combination of sharing and separating the parts"* (Fox & Faver, 1982:336).

Of course, a precondition for working collaboratively is that the people concerned can broadly agree how to tackle the problem at hand:

> *"In taxonomy it's virtually impossible to work with other people - like judges, taxonomists can give opinions but they don't give joint opinions"* (anonymous informant quoted in Becher, 1989:98).

Similar considerations may well apply to questions of information policy, which by their very nature are often of a highly political and sensitive nature.

Experimental findings

The incidence of multiple authorship in the test collection is indeed low, as the distribution in Table 6.9 shows. The low incidence of multiple authorship in the information policy bibliography (15.6 per cent) is broadly similar to patterns that have been observed in the general library and information science literature[34,35].

[34] In a study of the library and information science literature, Bottle & Efthimiadis (1984) noted a range of authorship patterns from relevant abstracting and indexing services for 1983. Their estimates of multiple authorship range from 19.1 per cent *(Computer & Control Abstracts)* to 36.6 per cent *(Information Science Abstracts)* to with an overall mean of 30.8 per cent. The work of Bottle & Efthimiadis was not restricted to the journal literature, nor is it comparable in terms of its time frame with the results reported here.
[35] Raptis (1992), in a study of 39 British and American library science journals (1950-75), found that 13.6 per cent of articles were multiple-authored.

Table 6.9: Distribution of multiple authorship

Distribution of co-authors	Number of articles, *A*
0	630
1	106
2	25
3 or more	10
Total articles	*771*

In recent decades the incidence of multiple authorship has increased significantly across virtually all disciplines which have been investigated. According to figures released by the *Institute for Scientific Information* (cited in Cronin, Davenport & Martinson, 1997), the average number of authors per article in *Science Citation Index* has increased from 1.84 (1966) to 3.67 (1995). The equivalent figures for the *Social Science Citation Index* are 1.15 and 1.74.

Information policy seems to be an exception to this trend: the mean number of authors per paper has not changed significantly over the period 1972-96 (Table 6.10):

Table 6.10: Author collaboration over time

Period	*n*	Mean authors per article
1972-76	41	1.10
1977-81	100	1.24
1982-86	163	1.20
1987-91	197	1.28
1992-1996	270	1.29
1972-1996	*771*	*1.22*

$p > 0.4$

The overall mean is quite low compared to the *SSCI* population as a whole and this may possibly be indicative of the weak institutionalisation of the field of information policy (see Section 6.4.3). It may also be a reflection of the often speculative and philosophical nature of the subject content.

6.4.3 Patterns of corporate authorship

One way of looking at disciplines or fields of study is to look not at the activities of individual authors but at the institutional arrangements which support and encourage their research. There may be doubts, for example, whether statistics is now sufficiently separate from its parent discipline, mathematics, to constitute a discipline in its own right? The answer will depend, in part, on the extent to which institutions recognise the hiving off in terms of their organisational structures and, in part, on the degree to which a free-standing international community has emerged with its own professional associations and specialist journals. The degree of institutionalisation of a field of studies is thus an important indicator of its maturity and status. Intense debates often surround the establishment of new forms of institutionalised knowledge production (such as queer studies, peace studies or parapsychology), whose intellectual validity may well find itself under attack from established academic opinion.

The degree to which knowledge production is concentrated in an institutional sense is also an issue which attracts policy attention. The concept of research selectivity, for example, seeks to shape academic activity in order to maximise administrative efficiency, economies of scale and optimise the division of labour (Hicks & Katz, 1997). This is achieved through systematic policies which encourage the concentration of research activities into a smaller number of more specialised units. As discussed in Chapter 2, there are other forces which tend to disperse rather than concentrate research efforts: possibly the two most important factors here are the diaspora created by the vastly-increased post-war production of PhDs and the emergence of new fora, outside traditional university settings, where 'academic' research activities take place.

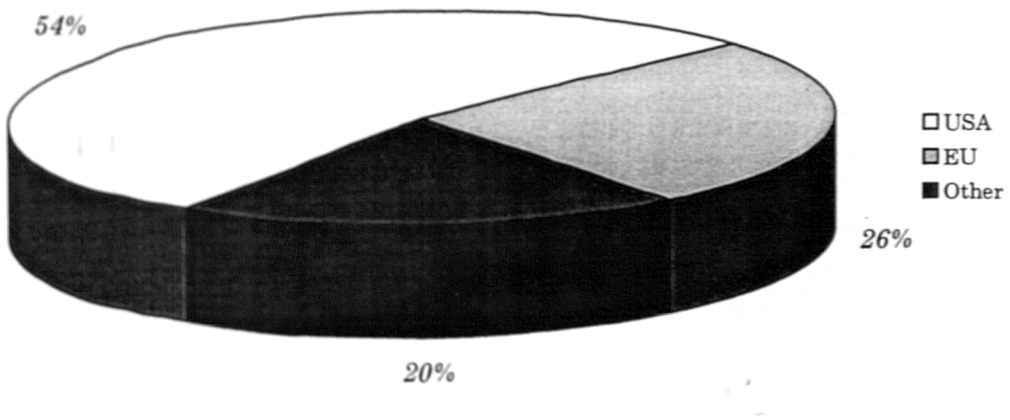
A comprehensive explanation of how fields of study become recognised and institutionalised lies well beyond the scope of this thesis and would require the development of a systematic framework for understanding the mechanisms which lead to the emergence of autonomous, self-generating research

structures[36]. The analysis of corporate authorship presented here has more modest aims: to develop a broad-brush picture of the degree to which institutions, rather than individuals, feature in the production of knowledge in information policy. This is achieved through the analysis of institutional names derived from the corporate affiliations field[37] in *SSCI*.

Geographical distribution

Figures 6.19 and 6.20 show the geographical distribution of unique corporate names (rather than individual articles) across the test collection, globally and for the countries of the European Union, respectively.

Figure 6.19: Geographical location of corporate authors

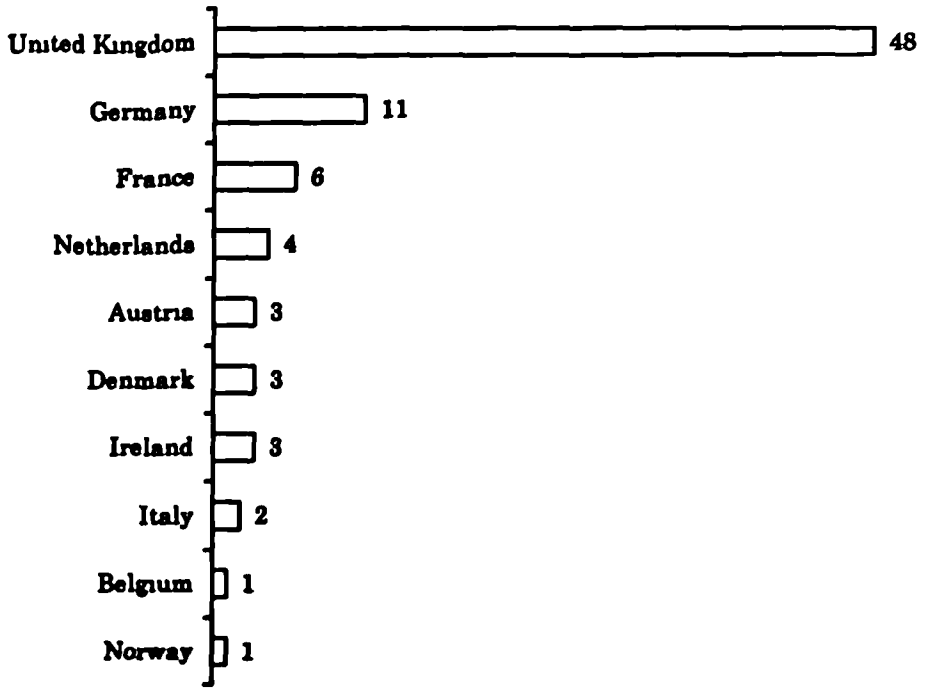


The data in Figure 6.19 shows a rather different pattern than that observed earlier for the geographical distribution of articles (see Figure 6.3, page 89), especially in relation to organisations outside Europe and the USA.

[36] The work of Becher & Kogan (1980) offers a good blueprint for how such a study might be carried out.

[37] *SSCI* provides corporate affiliations for every named author. The analyses which follow are based on the frequency with which unique corporate names occur; thus, collaborative papers may contribute two or more corporate names and are 'double-counted' here.

Figure 6.20: European corporate authors



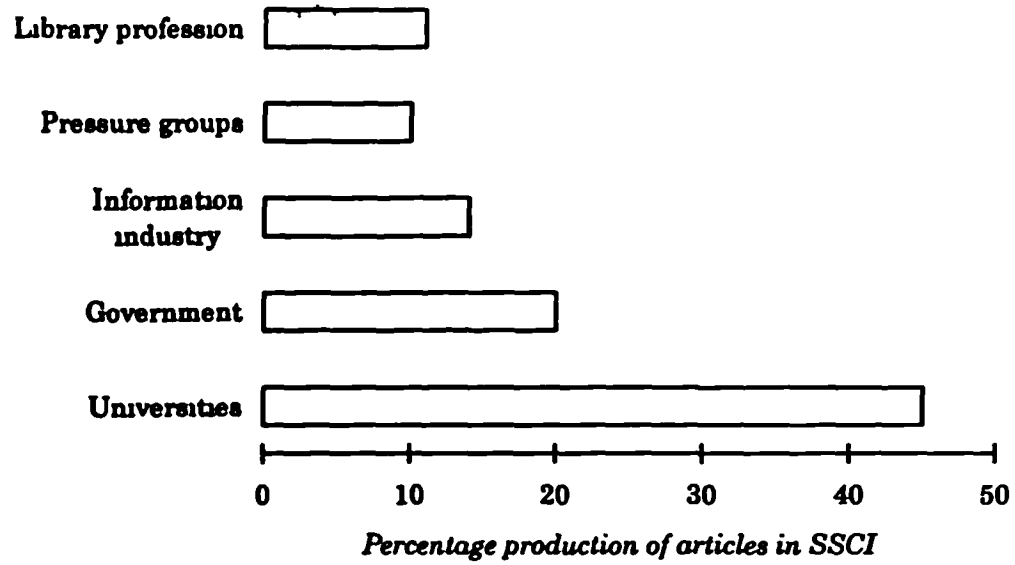
The breakdown of corporate addresses by EU Member State accords reasonably well with the findings of a survey by Stroetmann (1992). Stroetmann surveyed European organisations engaged in information research in the public, private and non-profit sectors in 1990. The survey identified 37 organisations (out of a total response of 86) which claimed special expertise in information policy research, the majority of which were located in the UK and then Germany. Stroetmann notes that it was evident from the questionnaires and the material sometimes accompanying them that a significant number of institutions had only recently been established and sometimes had just started research activities, or would do so in the near future. This was interpreted as a sign that "across Europe the information science field (in the broad sense defined above) is at last, even if slowly, becoming a recognised and established field of applied research and consultancy [which can] provide a service to those in need of solid research results and policy support" (Stroetmann, 1992:152). Might these same data also be taken as signs of incipient concentration?

Concentration and dispersal

A common theme in this thesis is that knowledge production in information policy is more heterogeneous than is commonly realised and that much research

takes place outside of the traditional academic setting. This view receives strong support from the evidence presented in Figure 6.21, which assigns unique corporate addresses to one of five sectors.

Figure 6.21: Dispersed knowledge production



In terms of the number of identifiable institutions, universities account for less than half of the total, suggesting that information policy research is indeed heterogenous and institutionally dispersed; an impression which is strongly reinforced when one examines the number and range of institutions listed in Appendix C.

Gibbons and others (1994) argue that socially dispersed knowledge production is one characteristic of a knowledge-based society. They also acknowledge that the forces that shape the relative concentration or dispersal of research activity are complex and that there are vectors acting in both directions. This raises some interesting questions. In information policy, are all sectors equally dispersed, or are some more concentrated than others? Over time, has information policy research become more, or less, associated with institutions rather than individuals?

The data in Table 6.11 overleaf show the top ranking information policy corporate authors (by article production) over the period 1972-1996. With a sole

exception (Commission of the European Communities) these are all US-based organisations.

Table 6.11: Top ranking corporate authors

Rank	Name of institution	Cum. % all articles
1	Library of Congress Congressional Research Service	4.54
2	Syracuse University School of Information Studies	6.87
3	Commission of the European Communities	8.69
4	Simmons College School of Library & Information Studies	10.38
5	University of Illinois	11.80
6	US Office of Management & Budget Watch	13.09
7	Rutgers State University	14.14
8 =	George Washington University	15.05
8 =	Suny University Albany	15.96
8 =	University of California Los Angeles	16.87
8 =	University of Michigan	17.78
8 =	University of Strathclyde	18.69
8 =	US Bureau of Census	19.60
8 =	US House of Representatives	20.51

Some further bibliometric evidence responding to these questions is presented in Table 6.12 overleaf. This analysis follows the example of Hicks & Katz (1996, 1997) who examined similar questions in relation to UK science policy.

Since it is difficult to directly compare two or more distributions and make meaningful comparisons, Hicks & Katz recommend the use of a derived measure, the Herfindahl Index. The Herfindhal Index (named after the late Orris Herfindahl) is a tool commonly used in economics to describe the equality or inequality of a distribution (such as *per capita* income). The general formula is: $H = (S_1)^2 + (S_2)^2 + ... + (S_n)^2$, where S_1 through S_n are the shares (totalling 100 per cent) of entities 1 through n. The index is often used as a measure of the competitiveness of a market or industry: a market comprising two firms with a market share of 95 and 5 per cent respectively (i.e. highly concentrated) would have an index value, H of 0.905. On the other hand, in a more competitive marketplace, with ten firms sharing the market equally, H falls to 0.1[38].

[38] The US Justice Department uses the Index to determine whether a merger would illegally restrain competition. In their guidelines, a merger would not be challenged if the resulting Herfindahl Index was less than 0.1 (Ruffin & Gregory, 1990).

Table 6.12: Concentration indicators

	Institutions	Articles	Herfindahl index	Sector size equivalent[39]
By sector				
Education	170	373	0.0122	82
Government	63	161	0.0702	14
NGOs	12	36	0.0201	50
Non-profit	51	78	0.0276	36
Industry	45	52	0.0296	34
By time period				
Before 1985	86	229	0.0036	278
Since 1985	255	451	0.0014	714
All	*341*	*680*	*0.0082*	*122*

A number of conclusions can be drawn from these results. The first and most obvious conclusion is that information policy research is very thinly dispersed in institutional terms. Of the five sectors, government is the most concentrated, education the most dispersed. Surprisingly, the field seems to have become more dispersed since 1985.

To put these figures into context, Hicks and Katz (1997) examined a sample of records from the *Science Citation Index* representing a broad cross-section of UK scientific research over the period 1981-1994 (Hicks & Katz, 1997). Their findings show that the UK scientific system is much less dispersed than information policy research. The respective index values for scientific research and information policy research are, for education 0.038 (0.012) and government 0.165 (0.070).

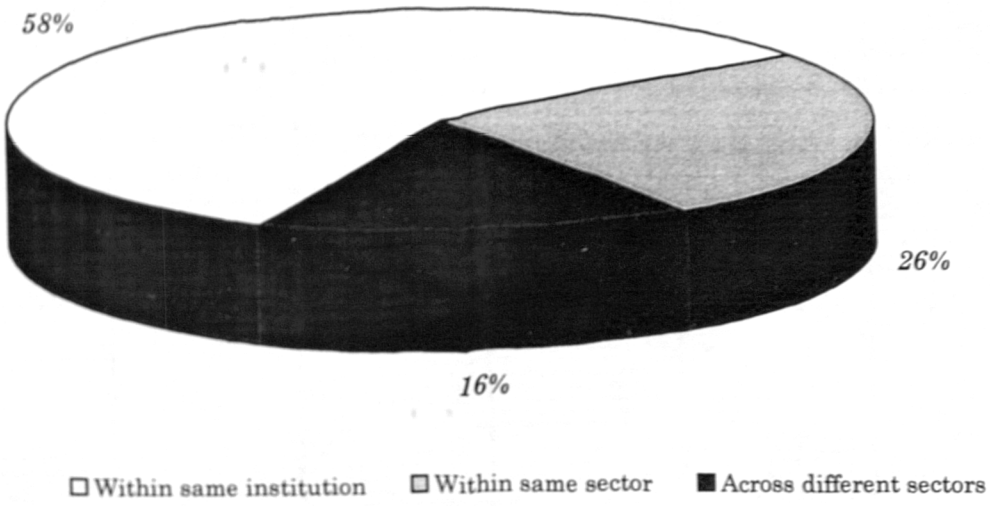
Collaboration between institutions

The overall pattern of collaboration between institutions (Figure 6.22) shows that collaborative activity tends to takes place within the same institution, rather than cutting across institutions or sectors. When these data were

[39] **Sector size equivalent.** In a sector of size n, concentration reaches a maximum when all institutions publish the same number of papers. At this minimum, the Herfindahl Index would equal $1/n$ where n is the number of institutions. By reversing this logic, each sector can be seen to be as concentrated as a sector of $1/H$ number of equal sized institutions.

disaggregated into two time periods, 1975-1984 and 1985-1996, no significant differences were noted.

Figure 6.22: Collaboration between institutions



6.5 Conclusions

In many ways, the profile of the information policy test collection which emerges from this study challenges some of the assumptions about the behaviour of social science literatures which are commonplace in writings on bibliometrics.

The pattern of growth in the test collection is not unusual. It conforms to a power model and appears to be doubling every six years. This is, however, a more rapid pattern of growth than that of the *Social Science Citation Index* as a whole, with the implication that information policy is commanding greater attention from authors and editors alike.

In Price's typology, the test collection would provide a very good example of a highly immediate or 'ephemeral' literature, of the kind commonly observed in the natural sciences. Citation practices in information policy draw more heavily on recent material than might be anticipated from Price's simple probabilistic model, the relatively small size of the archive notwithstanding. This finding is given added weight by the rapid ageing of information policy citations (with a

half-life of only three years), another indicator which might be held to be more typical of the natural sciences (see Table 6.13 below).

Table 6.13: Comparison of median citation ages in other fields

Field	Median citation age (yrs)
Rowlands' test collection	***3.0***
Metallurgical engineering	3.9
Chemical engineering	4.8
Genetics	5.0
Information systems	5.0
Physics	5.2
Mechanical engineering	5.2
Desalination	5.6
Chemistry	8.1
Archaeology	9.5
Botany	10.0
Mathematics	10.5
Geology	11.8
Music education	12.5
Music theory	12.5
Biblical criticism	21.6

(adapted from Cunningham & Bocock, 1995).

Intriguingly, there are hints in the literature that high levels of immediacy may be encountered in disciplinary areas that are undergoing revolutionary, paradigmatic change (Cozzens, 1985).

As might be expected for a bibliography representing a 'word-specialty', articles on information policy exhibit a spread across a large number of journal titles. It would be very tempting to characterise the information policy literature as showing a high degree of documentary scatter.

Table 6.14: Comparison of article/journal density with other studies

Bibliography	Source	Density
Lubrication	Bradford (1934)	2.41
History of psychology	Coleman (1993)	3.86
Mast cell	Goffman & Warren (1969)	4.05
Geophysics	Bradford (1934)	4.09
Articles citing Kuhn	Coleman (1993)	4.20
Rowlands' test collection		***4.26***
Schistosomiasis	Goffman & Warren (1969)	5.70
Human eyeblink conditioning	Coleman (1993)	6.76
Pavlovian conditioning	Coleman (1993)	8.77

(adapted from Coleman, 1993).

Coleman (1993) argues that article/journal densities offer a reasonably good indication of documentary scatter: the lower the density, the more thinly spread the distribution of articles across journal titles. The comparative data in Table 6.14 above suggest that the information policy bibliography is actually *less* thinly spread than such classic literatures as Goffman & Warren's mast cell or Bradford's lubrication articles. No conclusion is suggested here, merely the observation that the test collection does not appear to be unusually scattered. In this context, it is interesting that none of the core journal titles conform with the notion of a general purpose 'Journal of Information Policy'. Some are very general *(e.g. Aslib Proceedings, Proceedings of ASIS),* while others are more obviously policy-oriented *(e.g. Telecommunications Policy, Government Information Quarterly)* but problem- or sector-specific rather than field-specific.

Another striking finding is the very low incidence of collaborative authoring. The mean number of authors per paper is significantly lower than the corresponding figure for *Social Science Citation Index* as a whole (1.22 and 1.74 respectively, for 1995). In this respect, authoring behaviour more closely resembles that of scholars in the arts and humanities. Unlike many other social science disciplines which have seen an increase in collaborative authorship over the past 20 years, the low levels encountered in the information policy literature appear to be a structural feature. It is difficult to interpret this finding since a number of factors may be influential: the weakly institutionalised nature of the field within higher education, perhaps, or the lack of research funding, or it may simply be a distinctive feature of cognitive style.

Chapter 7: Cluster analysis

> *"As we have no written pedigrees, we are forced to trace community of descent by resemblances of any kind ... we care not how trifling a character may be —if it prevail throughout many species, especially those having very different habits of life, it assumes high value"*[40]

7.1 Introduction

The previous Chapter examined the distribution of various bibliometric elements across the test collection, one or two elements at a time. The work presented here extends that analysis by considering many elements *at the same time* using multivariate techniques.

The objective of this Chapter is to respond to an important bibliometric question: is the test collection homogenous, or is there any underlying natural structure? Put another way, is the test collection *(B)* best thought of as a single bibliography or as a series of joint bibliographies, B_1, B_2 ... B_n, possibly sharing common journals but not common papers? How far, in terms of the internal structure of the test collection, might any of these subsets diverge from that of the parent set? These are important questions, given the aim of this thesis which is to map the intellectual, organisational and social topography of information policy.

In this experiment, the underlying structure of the document test collection was explored using agglomerative hierarchical clustering techniques. Cluster analysis has been widely used in disciplines as varied as the social sciences, psychology and electrical engineering to classify objects into categories. Clustering techniques are particularly appropriate where little is known in

[40] Charles Darwin writing on the "mutual affinities of organic beings" in *The Origin of Species* (1859).

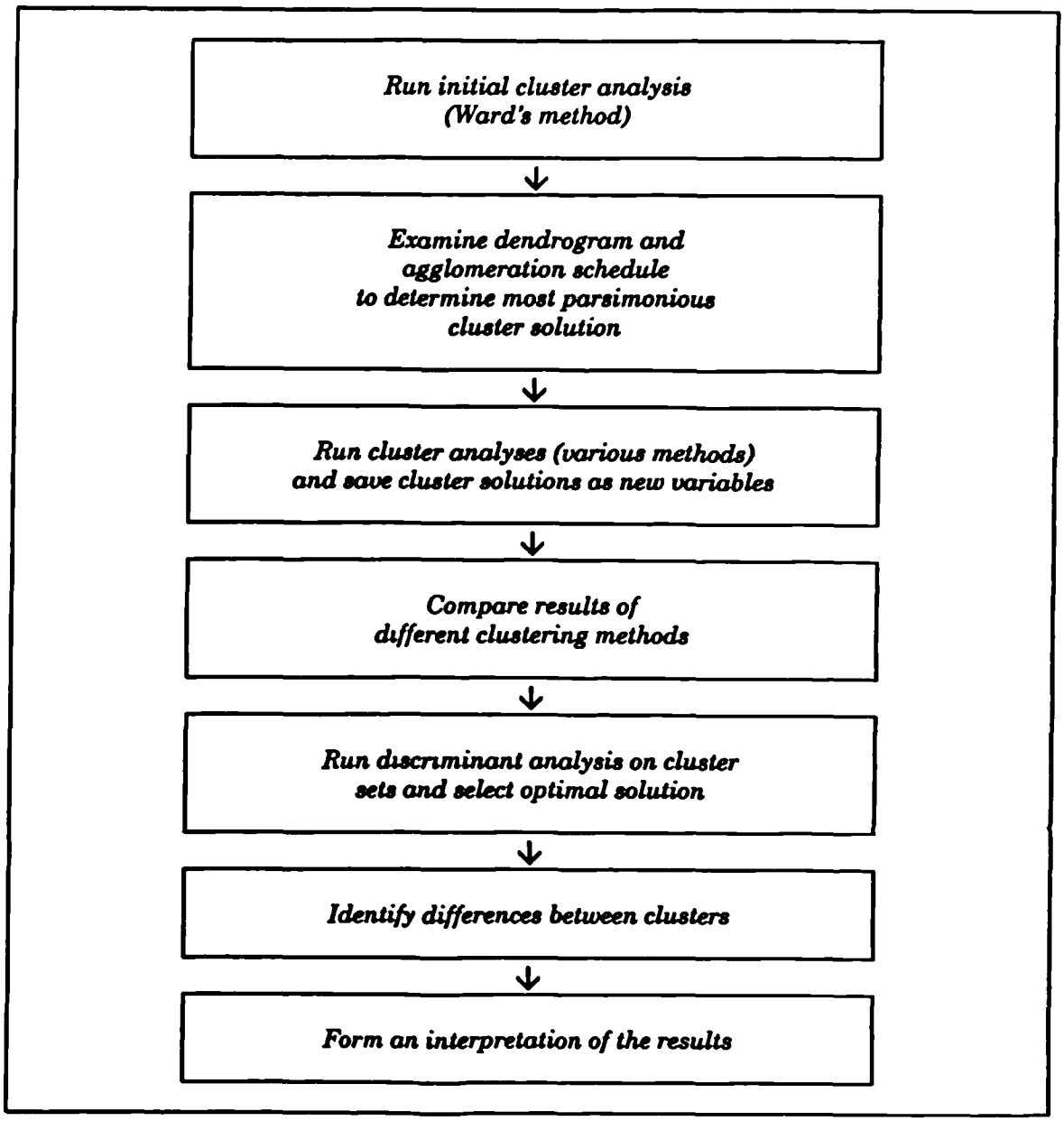
advance of the underlying data structure and where a typological interpretation is needed. Clustering methods are used to uncover natural groupings or types, yet unlike other modes for classification (e.g. supervised learning in pattern recognition studies) they do not require the use of categorical labels until the interpretation of the clusters is complete. It is not necessary to know either the criteria for group membership or even the number of groups *a priori*. It follows that cluster analysis is highly exploratory and heuristic in nature and it needs, ideally, to be supplemented by other methods to determine the validity (if any) of the resulting clusters. An analogy between this research approach and the work of naturalists in the eighteenth and early nineteenth centuries might be drawn. The foundations of biological taxonomy were established by carefully observing the similarities and differences between plants and between animals. From this systematic observation gradually emerged taxonomic structures which showed, for example, that hedgehogs and dolphins are more fundamentally similar to one another than are hedgehogs and sea urchins (despite the obvious fact that both are round and spiny, while dolphins are smooth and elongated). In the days before the emergence of Darwinian evolutionary theory, population genetics, and the elucidation of the structure of DNA, these were still useful insights and they provided a platform for later and more profound understanding.

7.2 Research design

The methodology for this experiment is summarised in Figure 7.1 overleaf.

The first decision to be made before the experiment could be conducted was which variables should be included; if important variables are excluded, poor or misleading findings may result. It is also important to choose variables that might be useful in developing a meaningful interpretation of the results; in cluster analysis, the initial choice of variables determines the characteristics that can be used to identify subgroups.

Figure 7.1: Research design



The twenty two variables selected for the experiment are shown in Figure 7.2 on the next page. They represent bibliometric and other indicators which are more descriptive of content. Each of the variables was taken (or derived) from the record structure of the *Social Science Citation Index*, with the exception of the information policy subject headings (SSBJ, SBJC) and scale categories (SCAL) discussed in Chapter 3.

Figure 7.2: Variables included in the cluster analysis

ACAD	*First author: academic or practitioner*
AUTH	*Number of authors*
BRDA	*Bradford zone of author productivity*
BRDJ	*Bradford zone of journal productivity*
GEOA	*First author: country*
GEOJ	*Journal: broad geographic region*
INDX	*Price's index*
INST	*Type of institution (first corporate author)*
JCAT	*ISI journal category*
LANG	*Language*
LISS	*Journal: LIS or other title*
RATE	*Mean annual citation rate*
RNKA	*Rank order of author productivity*
RNKJ	*Rank order of journal productivity*
SBJC	*Narrow subject heading*
SCAL	*Information policy scale*
SGEO	*First author: broad geographic region*
SSBJ	*Broad subject heading*
TIME	*Published before or after 1986*
TYPE	*Research or opinion paper*
WIND	*Citations present or absent*
YEAR	*Year of publication*

In the initial cluster run, using Ward's method (Sneath & Sokal, 1973), the variables were normalised using Z-scores and a dissimilarity matrix of simple Euclidean distances was created. Visual inspection of the resulting dendrogram and agglomeration schedule suggested that a 6-cluster solution offered the most parsimonious interpretation of the data. The data were then clustered again with a 6-cluster solution specified in advance so that a new *SPSS* variable ('CWA6') could be created, reflecting the assignment of individual cases to the final clusters. At this stage, two other algorithms, group average (within groups) and single linkage (or nearest neighbour), were applied so that comparisons could be drawn between the various clustering methods. Comparative studies have shown that no one single clustering algorithm is best in every situation and that often, it is the mathematically respectable single-linkage which is the least successful in many applications (Milligan, 1980).

Cluster analysis methods do not lead one to a purely objective and stable classification, in spite of the optimism voiced by workers in numerical taxonomy (see, Sneath & Sokal, 1973, for instance). They are better regarded as tools for exploring data rather than for the production of formal taxonomies. By creating

new *SPSS* variables representing group memberships, it then became possible to use other statistical techniques, notably discriminant analysis and cross-tabulation, to explore the factors which were instrumental in partitioning of the test collection. The outputs of these techniques were finally used to construct an interpretation of the six-cluster solution produced by Ward's method.

7.3 Cluster validity

The ultimate aim of clustering is to obtain an optimal partitioning of the basic object set into subsets which are compact and mutually isolated. Clustering algorithms will, however, inevitably group objects, irrespective of the presence or absence of any natural structure in the data.

Graphical aids, such as dendrograms and icicle plots are available which portray the multidimensional data generated in clustering procedures and facilitate informal clustering by eye. Purely qualitative interpretations of clusterings are not very rigorous, however, and are unlikely to be able to resolve the difference between 'real' clusters and those which arise as artifacts of the clustering methodology. Tools are clearly needed to support the search for objective meaning or 'cluster validity', especially in applications where little prior categorical information about the data is available. A number of approaches to the problem have been discussed in the statistical literature, but few generally applicable tests have been developed (Rowlands, 1983). Broadly speaking, measures of cluster validity address themselves to three fundamental questions:

- is there any intrinsic (i.e. non-random) structure in the data to be clustered?
- how well does the clustering recover the 'true' overall data structure?
- how 'good' are the individual partitions?

In the next sections, some answers to these questions are attempted, admittedly in a somewhat superficial way, given the current lack of generally agreed practice for undertaking cluster validity studies. Dubes & Jain advocate the use

of validity studies as a tool for comparing the efficacy of different algorithms and advise the potential user of cluster analysis to "... apply several clustering approaches and check for common clusters instead of searching for a technical measure of validity for an individual clustering" (Dubes & Jain, 1979: 254).

7.3.1 Comparison between different clustering methods

It has already been noted that all clustering methods will continue to partition a set of objects until exhaustion, regardless of the presence of any underlying structure. One measure of whether the clusters produced by a particular method are arbitrary or whether they reflect some kind of structure is to see whether different methods tend to assign the same cases to the same clusters. Intuitively, if two different clustering methods assign cases to clusters on a completely arbitrary basis, one would expect there to be relatively little overlap between the resulting clusters. If the two methods are assigning cases at random, it should be possible to detect this by simply applying a non-parametric test such as chi-square[41]. Table 7.1 shows the overlaps of cluster membership produced by clustering the test collection using the three different methods. The null hypothesis asserted is that each method allocates cases to clusters in a random fashion; thus, the *expected* value for each cell is 771/36 = 21.4. Cases allocated to the same cluster by both methods are indicated in bold type.

Table 7.1: Case assignments: comparison of three clustering methods

Ward's method vs group average[42]

	GA 1	*GA 2*	*GA 3*	*GA 4*	*GA 5*	*GA 6*	*All*
WM 1	**175**	49	2	11		2	239
WM 2	13	**35**		8		2	58
WM 3	1		**72**	3			76
WM 4	3	9		**119**			131
WM 5	24	1	3	15	**33**	79	155
WM 6	40			20		**52**	112
All	256	94	77	176	33	135	771

Overlap: 486 (63.0 per cent).

[41] I am indebted to Professor Steven Robertson for his advice on this issue.
[42] Chi-square = 2354.9, d.f. = 25, $p < 0.001$

Table 7.1: Case assignments: comparison of three clustering methods *continued*

Ward's method vs single-linkage[43]

	SL 1	*SL 2*	*SL 3*	*SL 4*	*SL 5*	*SL 6*	*All*
WM 1	236	1		2			239
WM 2	38		18		1	1	58
WM 3	76						76
WM 4	131						131
WM 5	155						155
WM 6	112						112
All	748	1	18	2	1	1	771

Overlap: 236 (30.6 per cent).

Single-linkage vs group average[44]

	GA 1	*GA 2*	*GA 3*	*GA 4*	*GA 5*	*GA 6*	*All*
SL 1	243	94	77	167	33	134	748
SL 2	1						1
SL 3	10			7		1	18
SL 4	1			1			2
SL 5				1			1
SL 6	1						1
All	256	94	77	176	33	135	771

Overlap. 243 (31 5 per cent).

All three pairs of methods produce results which require the firm rejection of the null hypothesis. This finding needs careful interpretation, however: it should not be taken as proof that there is any natural structure in the data. Nor, if the test collection is highly structured, does it mean that the resulting clusters accurately reflect that structure. It merely shows that the three methods did not cluster the documents at random (perhaps because all three algorithms resulted in similarly distorted and artificial results).

Looking at Table 7.1 it is interesting to note that the three methods produce very different clustering patterns, albeit with significant overlap. The single-linkage method allocated 97 per cent of the documents to a single cluster. Should this be taken to mean that the test collection is homogenous? On the other hand, the other methods produced a much more even distribution of documents across clusters (especially Ward's method). Does this suggest that the test collection is better regarded as a series of joint bibliographies rather than a homogenous collection? With these thoughts in mind, the next phase of the experiment

[43] Chi-square = 4694.6, d.f. = 25, $p < 0.001$
[44] Chi-square = 4878.1, d.f. = 25, $p < 0.001$

explored the clusters formed by Ward's method to see whether a viable interpretation could be made of the group memberships. Before that, however, a further test of the sensitivity and stability of the Ward clusters was made.

7.3.2 Cluster sensitivity and stability

How stable are the Ward clusters? If a clustering algorithm is to provide an appropriate summary of the inherent structure of a data set, the hierarchies generated should be tolerant of minor perturbations, such as might be caused by omitting small groups of objects from the classification or by adding noise to each of the objects (Milligan, 1978).

The data in Table 7.3 overleaf show the effects of removing a relatively small number of documents from the test collection. In each case, 40 randomly selected documents (approximately 5 per cent) were removed and the remaining documents reclustered. The null hypothesis this time was that the removal of these documents would make no difference to the final outcome: in other words, the remaining documents would all be allocated to the same cluster as before. For the sake of clarity, this concept is represented schematically as Table 7.2. This shows the overlaps between the original clusters (WM_{ALL}) and the randomly selected subsets (WM_{RAND}) that would be anticipated if the null hypothesis holds. The overlaps are expressed as percentages, 100 or 0, for diagonal and off-diagonal cells respectively.

Table 7.2: Sensitivity and stability of Ward clusters: 'null hypothesis'

	$WM\ 1_{ALL}$	$WM\ 2_{ALL}$	$WM\ 3_{ALL}$	$WM\ 4_{ALL}$	$WM\ 5_{ALL}$	$WM\ 6_{ALL}$
$WM\ 1_{RAND}$	100%	0%	0%	0%	0%	0%
$WM\ 2_{RAND}$	0%	100%	0%	0%	0%	0%
$WM\ 3_{RAND}$	0%	0%	100%	0%	0%	0%
$WM\ 4_{RAND}$	0%	0%	0%	100%	0%	0%
$WM\ 5_{RAND}$	0%	0%	0%	0%	100%	0%
$WM\ 6_{RAND}$	0%	0%	0%	0%	0%	100%

It was of course not possible to apply the chi-square test in this instance, since this would imply division by zero. However, the results of ten iterations of this rather informal test are shown in Table 7.3.

Table 7.3: Effects of data perturbation on Ward cluster membership[a]

Iteration	*WM 1RAND*	*WM 2RAND*	*WM3RAND*	*WM4RAND*	*WM5RAND*	*WM6RAND*	*ALL*
1	89.5	84.8	87.3	99.6	80.4	91.9	89.3
2	90.0	73.0	88.9	98.7	66.4	93.0	86.6
3	58.4	23.7	97.4	96.1	90.0	0.0	64.1
4	99.1	96.6	94.9	99.6	62.7	83.8	89.0
5	59.5	43.2	76.1	76.8	72.8	54.4	64.8
6	18.9	1.8	95.8	86.3	71.4	98.1	58.7
7	83.7	92.7	94.4	93.6	58.4	84.0	82.1
8	13.2	0.0	91.7	83.1	63.3	85.8	52.4
9	40.9	0.0	94.4	100.0	70.7	97.2	67.4
10	6.9	0.0	95.8	80.6	70.7	89.6	57.9
MEAN	56.0	41.6	91.7	91.4	70.7	77.8	71.2

[a]Percentages.

These data suggest that the 6-cluster solution provided by Ward's method is rather unstable, although clusters 3 and 4 are relatively unaffected by data perturbation. However, at this point, it well worth bearing in mind the view expressed by Dubes & Jain: "... one should not expect a single statistic to serve as a panacea for all problems in cluster validity ... too many factors are involved to expect a single statistic to cover the validity of clusters even for a single class of problems" (Dubes & Jain, 1979:253). Another, possibly more productive approach to the problem of establishing cluster validity, the one adopted in the remaining sections, is to use highly informal indices of authenticity based upon qualitative interpretation and the application of standard statistical procedures[45].

7.3.3 Discriminant analysis

Cluster analysis and discriminant analysis are closely related techniques. Both classify objects into categories, although in the case of discriminant analysis it is essential to know group membership in advance so that the classification rules can be derived. In this phase of the study, the six clusters created by Ward's method were subjected to a discriminant analysis[46], utilising the same variables that were used to create the clusters (Figure 7.2, p.128). The discriminant

[45] However, some authors (e.g. Turner, 1969; and Gnanadesikan and others, 1977) argue that the use of these standard statistical procedures are at best dubious, at worst comprehensively inadequate, even when the usual assumptions of multivariate analysis such as equality of scatter and Gaussian distributions in all clusters have been made.

[46] The variables were entered stepwise using Wilk's Lambda method (Norušis, 1994) and thresholds of F for entry and removal of 3.84 and 2.71 respectively.

analysis correctly predicted the assignment of individual articles to one of the six Ward clusters in 86.9 per cent of cases.

One of the most useful outputs of discriminant analysis in *SPSS* is the structure matrix (Table 7.4). This shows the pooled within groups correlations between the original variables and a smaller number of discriminant factors (DV_{1-5}) which are labelled impressionistically for ease of interpretation. Values greater than 0.35 are indicated in bold type.

Table 7.4: Discriminant analysis: structure matrix

	1	2	3	4	5
DV_1: Scholarliness					
Windsor coefficient	**.925**[b]	.143	-.255	-.029	-.114
Price Index[a]	**.306**[b]	.093	-.106	.002	-.090
DV_2: Nationality					
Author country	-.118	**.439**[b]	.141	-.298	.264
Author region	-.107	**.398**[b]	.134	-.257	.279
Language	.003	**.385**[b]	.092	-.331	.088
Broad subject[a]	.008	-.068[b]	.022	-.028	-.004
Narrow subject[a]	.003	-.062[b]	.018	-.027	.007
DV_3: Journal category					
LIS journal	.265	-.355	**.639**[b]	-.149	.350
Journal category	.218	-.312	**.557**[b]	-.149	.471
Rank order: journal	.107	-.106	**.399**[b]	-.094	.221
Bradford journal zone[a]	.096	-.041	**.382**[b]	-.095	.166
Research papers[a]	-.063	-.067	.230[b]	.068	.033
Mean ACR	.033	-.025	-.115[b]	.072	.043
DV_4: Productivity					
Rank order author	.014	.324	.425	**.722**[b]	.307
Bradford author zone[a]	.005	.343	.379	**.662**[b]	.289
Number of authors	.039	.066	-.042	.166[b]	.140
DV_5: Age					
Time period	-.010	-.028	-.511	-.046	**.776**[b]
Year of publication[a]	-.066	.013	-.427	-.043	**.550**[b]
Journal title origin	-.032	.129	-.015	-.023	.268[b]
Academic / practitioner[a]	-.055	.033	.049	.040	-.209[b]
Info policy scale[a]	-.035	-.038	.008	.110	-.120[b]
Institutional affiliation[a]	-.045	-.063	.023	.048	-.108[b]

[a]This variable was not used in the analysis.
[b]Largest correlation between each variable and any discriminant function.

The value of the structure matrix is that it enables an interpretation of the factors which account for most of the variance in the data to be made. In this case, five discriminant factors (DV) are indicated which, together account for 100 per cent of the variance. The five factors are summarised on the next page as Table 7.5.

Table 7.5: Discriminant analysis: Eigenvalues

Function	Discriminant factor	Eigenvalue	Percent of variance	Cumulative percent	Canonical correlation
1	Scholarliness	3.160	33.3	33.3	.872
2	Nationality	2.314	24.4	57.6	.836
3	Journal category	2.097	22.1	79.7	.823
4	Productivity	1.092	11.5	91.2	.722
5	Age	.839	8.8	100.0	.675

It is interesting to note that not all the variables played a role in constructing the final discriminant analysis: information policy topics, institutional affiliation and functional role (academic, practitioner) were not retained. The most important factors in predicting cluster membership were scholarliness, nationality and journal category. These factors account for nearly 80 per cent of the cumulative variance.

7.4 Cluster interpretation

The results of the discriminant analysis suggest that Ward's method may indeed be uncovering real structure in the test collection: it is possible to predict Ward cluster membership 17 times out of 20 based just on a knowledge of the five discriminant factors mentioned above.

A convincing interpretation, however, requires that the six Ward clusters are reasonably compact and mutually isolated. Table 7.6 overleaf assists the final interpretation by cross-tabulating the most highly correlated variable associated with each of the five discriminant factors by cluster membership. It should be read together with Tables 7.7 (which summarises the modal or mean values for each variable) and 7.8 (which shows, for each cluster, the six `most typical' articles for each cluster as determined by those closest to the cluster centroids). A final summary interpretation of the Ward clusters is presented as Figure 7.3 and is discussed more fully in Chapter 10. The results of the Ward clustering method yielded groups for which it was surprisingly easy to construct a coherent and persuasive interpretation; clusters 2, 5 and 6 are particularly sharply-defined. The clusters seem intuitively to be more internally consistent than is the test collection as a whole. On the evidence of this cluster analysis, it would

appear that the bibliography is structured and that it is therefore appropriate to regard it as a hybrid entity comprising a number of joint bibliographies.

Table 7.6: Ward clusters by discriminant factors

Scholarliness factor: Ward clusters by citations present or absent[47]

Cluster:	*1*	*2*	*3*	*4*	*5*	*6*
No citations	3	3	75	0	17	0
	-6.4	*-1.8*	*23.7*	*-4.8*	*-0.7*	*-4.4*
Citations	236	55	1	131	138	112
	6.4	*1.8*	*-23.7*	*4.8*	*0.7*	*4.4*

Nationality factor: Ward clusters by author region[48]

Cluster:	*1*	*2*	*3*	*4*	*5*	*6*
North America	175	4	40	119	119	105
	0.1	*-11.8*	*-4.2*	*5.1*	*1.2*	*5.4*
European Union	60	17	31	12	22	7
	2.7	*2.0*	*5.0*	*-3.2*	*-1.8*	*-3.8*
Other OECD	3	16	2	0	11	0
	-2.7	*9.3*	*-0.7*	*-2.6*	*2.1*	*-2.4*
Developing countries	1	21	3	0	3	0
	-3.2	*13.8*	*0.2*	*-2.4*	*-1.3*	*-2.2*

Journal factor: Ward clusters by LIS or other journal title[49]

Cluster:	*1*	*2*	*3*	*4*	*5*	*6*
LIS journal	229	50	74	119	8	60
	10.5	*2.8*	*5.5*	*5.7*	*-19.7*	*-4.1*
Not LIS journal	10	8	2	12	147	52
	-10.5	*-2.8*	*-5.5*	*-5.7*	*19.7*	*4.1*

Author productivity factor: Ward clusters by author Bradford zone[50]

Cluster:	*1*	*2*	*3*	*4*	*5*	*6*
Bradford zone 1	6	1	9	74	12	6
	-6.2	*-2.8*	*-0.6*	*15.4*	*-2.5*	*-2.9*
Bradford zone 2	9	5	12	57	9	15
	-5.4	*-1.2*	*0.5*	*10.8*	*-3.3*	*-0.2*
Bradford zones 3-7	224	52	55	0	134	91
	9.0	*3.1*	*0.1*	*-20.2*	*4.5*	*2.3*

Age factor: Ward clusters by time span[51]

Cluster:	*1*	*2*	*3*	*4*	*5*	*6*
Before 1985	13	21	29	2	42	112
	-9.5	*1.4*	*2.0*	*-7.5*	*-0.4*	*18.2*
Since 1985	226	37	47	129	113	0
	9.5	*-1.4*	*-2.0*	*7.5*	*0.4*	*-18.2*

[47] Chi-square = 573.4, d.f. = 5, $p < 0.001$; Cramer's V = 0.86, $p < 0.001$
[48] Chi-square = 377.7, d.f. = 15, $p < 0.001$; Cramer's V = 0.40, $p < 0.001$
[49] Chi-square = 462.3, d.f. = 5, $p < 0.001$; Cramer's V = 0.77, $p < 0.001$
[50] Chi-square = 435.4, d.f. = 10, $p < 0.001$; Cramer's V = 0.53, $p < 0.001$
[51] Chi-square = 396.2, d.f. = 5, $p < 0.001$; Cramer's V = 0.72, $p < 0.001$

Table 7.7: Ward clusters: independent variables (modal or mean values)

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
ACAD	Practitioner	Practitioner	Practitioner	Practitioner	Academic	Practitioner
AUTH	1.48	1 26	1.09	1 08	1.25	1.11
BRDA	Zones 3-7	Zones 3-7	Zones 3-7	Zone 1	Zones 3-7	Zones 3-7
BRDJ	Zone 2	Zone 4	Zone 3	Zone 1	Zone 4	Zone 5
GEOA	USA	Germany	USA	USA	USA	USA
GEOJ	North America	Europe	North America	North America	North America	North America
INDX	0.62	0.54	<0.01	0 63	0.48	0.61
INST	Academic department (non-LIS)	Academic department (LIS)	Government department or agency	Academic department (LIS)	Academic department (non-LIS)	Academic department (non-LIS)
JCAT	LIS	LIS	LIS	LIS	Law	LIS
LANG	English	English	English	English	English	English
LISS	Yes	Yes	Yes	Yes	No	Yes
RATE	0.33	0.07	0.13	0.30	0.19	0.14
RNKA	73	72	61	13	68	66
RNKJ	10	26	14	8	50	29
SBJC	Telecoms Policy	Libraries, Archives and Public Records	National and International Policies	IRM Policy and Practice	Freedom of Access to Information	Freedom of Access to Information
SCAL	National	National	National	National	National	National
SGEO	North America	European Union	North America	North America	North America	North America
SSBJ	Information Infrastructure Policies	Information Infrastructure Policies	General Articles on Information Policy	Information Management in Government	Information Access and Control	Information Access and Control
TIME	Since 1985	Since 1985	Since 1985	Since 1985	Since 1985	Before 1985
TYPE	Research	Opinion	Opinion	Research	Research	Opinion
WIND	Citations	Citations	No citations	Citations	Citations	Citations
YEAR[a]	5.5	9.0	10.0	6.5	7.5	16.5

[a]Median age of papers in mid-1996.

Table 7.8: Ward clusters: most typical documents

Cluster 1: Information infrastructure analysts

JE Clarke and others (1988). Federal information. foundation for national competitiveness. *Government Information Quarterly 5(4)* 353-365.
RK Anderson and others (1994). Developing a health information infrastructure for Arizona. *Bulletin of the MLA 82(4)* 396-400.
TA Byrd and others (1992). Experiences with information technology planning in State government. *Government Information Quarterly 9(2)* 135-150.
P Rafferty and others (1988). Corporate citizenship and monopolistic practices in the information industry: a study of IBM. *Aslib Proceedings 40(3)* 69-77.
MJA Trigo and others (1996). National information policy: the Portuguese national framework. *Journal of Information Science 22(3)* 219-227.
G McMurdo & E Simpson (1994). The White House files. *Journal of Information Science 20(5)* 306-313

Cluster 2: International library community

Y Sugimoto (1984). A comparative study of science and technology information policy in advanced countries. *Library and Information Science 22* 143-159.
Y Sugimoto (1985). A comparative study of information policies of OECD, EC, UNESCO. *Library and Information Science 22* 143-159.
M Bergmann (1987). Transborder data protection. *Nachrichten fur Dokumentation 38(5)* 293-296.
H Burkert (1990). The general right of access to government information: a regulatory task. *Nachrichten fur Dokumentation 41(2)* 103-108.
J Becker (1987). Transborder data flow between East and West: the situation of the Federal Republic of Germany. *Nachrichten fur Dokumentation 38(1)* 1-5.
L Bohmuller (1989). Cultural and library policy (1933-45): from the archives of the Jena University Library. *Zentralblatt fur Bibliothekswesen 103(5)* 205-211.

Cluster 3: Policy mandarins

RK Appleyard (1981). The role of the European Commission in the information environment. *Journal of Information Science 3(6)* 257-260.
T Riley (1981). Developments in freedom of information internationally. *Canadian Library Journal 38(3)* 137-141.
R Penn (1994). Australia. *Journal of Government Information 21(6)* 605-611.
A Missen (1986). Freedom of information: a progress report. *Government Information Quarterly 3(2)* 133-139.
RF de Bruine (1994). European Commission policy in the framework of the information services market. *Journal of Government Information 21(1)* 15-24.
CJ van Rosendaal (1984). European information policy situation. *Aslib Proceedings 36(1)* 15-23.

Cluster 4: Established LIS professionals

HC Relyea (1993). Security classification reform. *Government Information Quarterly 10(3)* 333-340.
HC Relyea (1994). The administration and operation of the Freedom of Information Act. *Government Information Quarterly 11(3)* 285-299.
HC Relyea (1994). A statutory mandate for security classification policy and procedure. *Government Information Quarterly 11(4)* 347-355.
HC Relyea (1989). Information policy, national security and the Reagan administration. *Government Information Quarterly 6(4)* 365-382.
HC Relyea (1985). Public access to Congressional records: present policy and reform considerations. *Government Information Quarterly 2(3)* 235-256.
MR Irwin (1987). National security and information technology: the new regulatory option. *Government Information Quarterly 4(4)* 359-369.

Cluster 5: Information regulation academics

W Kingston (1994). Compulsory licensing: an alternative to grants of monopoly with intellectual property. *Research Policy 23(6)* 661-672.
SE Goodman (1987). The information technologies and Soviet society: problems and prospects. *IEEE Transactions 17(4)* 529-552.
S Greenstein (1993). Markets, standards and the information infrastructure. *IEEE Micro 13(6)* 36-51.
E Scrivens (1986). Information technology and social policy. *Social Policy Administration 20(2)* 117-124.
RK Lindner (1993). Privatizing the production of knowledge: agricultural research. *Australian Journal of Agricultural Economics 37(3)* 205-225.
D Tsanacas (1985). The transborder data flow in the New World Information Order: privacy or control? *Review of Social Economy 43(3)* 357-370.

Cluster 6: Liberal lawyers

R Blasingame (1979). Public library and information policy. *Library Journal 104(16)* 1818-1822.
BO Slanker (1976). Public libraries and information industry. *Drexel Library Quarterly 12(1/2)* 139-148.
J Feeley (1980). Freedom of information: attitudes, rights, laws and policy. *Canadian Journal of Information Science 5(1)* 1-9.
WG Stiles (1980). Freedom of information and our state libraries. *Canadian Library Journal 37(6)* 391-395.
P Phillips (1980). Are trade secrets dead? The effects of the TSA and FoIA on trade secrets. *Journal of the Patent Office Society 62(11)* 652-677.
JVT Knoppers (1983). Transborder data flow issues and library information services. *Canadian Journal of Information Science 8(1)* 7-17.

Figure 7.3: Summary interpretation of clusters[52]

Cluster 1: 'Information infrastructure analysts'
Narrow subject: Information industry regulation (4.6)

Low productivity authors writing almost exclusively (95.8%) in the LIS journal literature. Broad range of interests and backgrounds, with an emphasis on information regulation and infrastructure. The youngest of the clusters: median age 5.5 years and very few papers before 1985; also the most highly collaborative. Highly immediate literature with a slight European bias.

Cluster 2: 'International library community'
Narrow subject: Libraries, archives and public records (4.6)

Highly international cluster with scarce North American representation (100% of foreign language articles are located here). Authors show a strong tendency to be LIS academics or library professionals writing opinion papers on practical library issues.

Cluster 3: 'Policy mandarins'
Narrow subject: National and international policies (4.6)

Mostly (90.4%) non-academic authors, often working in the government sector and writing opinion papers on national and international policy for LIS journals as sole authors. Mature literature with a slight European bias.

Cluster 4: 'Established LIS professionals'
Narrow subject: IRM policy and practice (3.5)

Mostly highly productive authors writing research papers, usually on their own, for core journals. Broad range of backgrounds but with a tendency to be associated with North American academic LIS departments or professional associations. Emphasis on information management issues in government.

Cluster 5: 'Information regulation academics'
Narrow subject: Copyright and IPRs (4.6)

A tight cluster of mainly US academics writing on freedom of information, copyright and intellectual property issues. Strongly represented in departments of law, politics or the social sciences. Rarely publish in LIS journals (5.1%).

Cluster 6: 'Liberal lawyers'
Narrow subject: Freedom of information (6.8)

A now extinct genre; no papers published since 1985. Legal practitioners and legal academics writing almost exclusively on freedom of information, confidentiality and privacy issues in the USA. Band-wagon effect?

[52] The figures in brackets are Pearson residuals. These express the differences between the observed and expected values (as in chi-square analysis) such that the differences have a mean of 0 and a standard deviation of 1. The narrow subject headings indicated are those with the highest residual value for each cluster.

Chapter 8: Structural bibliometric analysis

> ***"And with you there shall be a man of every tribe; every one head of the house of his fathers"***[53]

8.1 Introduction

In the previous two Chapters, a series of experiments was carried out with a view to identifying some of the key structural characteristics of the information policy serials literature, more specifically how certain bibliometric entities (authors, journal titles, research methodologies, institutional affiliations, etc.) were distributed across the population of articles and through time.

Many other studies into the structure of scientific disciplines and specialties have been based on the analysis of subject bibliographies. These bibliographies are however inevitably subject to bias. Small (1977), for example, argues that the difficulties inherent in making 'objective' relevance judgements are such that it is next to impossible for other researchers to replicate this type of study. One alternative to the subject-based approach is to employ citation linkages, at a given threshold level, to automatically generate sets of related documents. These document sets are more easily replicated and more objective than manually-created subject bibliographies, and they are much cheaper to produce. Document sets formed in this way are the raw material for cocitation studies, a branch of bibliometrics which is concerned with the mapping of scientific, scholarly and technical publications (White & McCain, 1989).

This Chapter presents the findings of an author cocitation study in the field of information policy. While still lying firmly in the bibliometric tradition, the approach in this Chapter differs fundamentally from what has gone before; some of the key differences are summarised in Table 8.1 overleaf.

---

[53] Old Testament, Book of Numbers 1:iv (King James version).

Table 8.1: Comparison of census and author cocitation studies

	Census studies (Chapters 3 and 4)	Cocitation studies (Chapter 5)
Production of bibliography	manual	automatic
Basis for production	subject-based	based on citation linkages
Inclusion criteria	relevance judgements	thresholds of 'citedness'
Methodological approach	'subjective'	'objective'
Unit of analysis	population of articles	authors' oeuvres
Research aim	to understand how certain entities are distributed across the whole population	to uncover aspects of the cognitive structure of the field of study

The reasons for undertaking an author cocitation study at this point are two-fold. Firstly, there was a need to address the problem raised by Small (1977) in relation to the subjectivity which inevitably informs the compilation of subject bibliographies. Author cocitation analysis (ACA) is less prone to this particular criticism, raising instead other, unrelated methodological concerns. Secondly, author cocitation analysis offers a powerful tool for understanding the social and intellectual structure of the information policy research community in a way that is not possible within the classical framework of simple bibliometrics.

The specific objectives of this Chapter are to:

- explore the community structure of the information policy domain
- identify relationships between different topics within information policy
- identify relationships between information policy and neighbouring disciplines

8.2 Cocitation studies

Cocitation studies are founded on the belief that it is possible to gain new insights into the structure of literatures by examining patterns of citation. Two key assumptions behind cocitation studies are (a) that the literature is the primary formal channel of communication among scientists and other scholars,

and (b) that citation patterns offer useful, if imperfect, indicators of scholarly activity and accomplishment. The primary use of cocitation techniques to date has been as a tool for studying the social and cognitive substructure of various disciplines and fields of study.

Published cocitation studies include investigations in both the natural and social sciences:

- collagen research (Small, 1977)
- particle physics (Sullivan, White & Barboni, 1977)
- information science (White & Griffith, 1981, 1982)
- technology transfer (Cottrill, Rogers & Mills, 1989)
- sociology of marriage and the family (Bayer, Smart & McLaughlin, 1990)
- organisational behaviour (Culnan and others, 1990)
- macroeconomics (McCain, 1990)
- research into decision support systems (Eom, 1996)
- artificial intelligence (van den Besselaar & Leydesdorff, 1996)
- scientific communication (Kärki, 1996).

Most published studies date have been based on data from *Science Citation Index (SCI)* or *Social Science Citation Index (SSCI)*, often using data supplied to order on magnetic tape or other carriers.

Four distinct types of cocitation study may be found in the literature (see, for example, Bellardo, 1980; White & Griffith, 1982; White & McCain, 1989; Osareh, 1996b):

- document cocitation analysis (DCA)
- journal-by-journal citation analysis
- country-by country citation analysis
- author cocitation analysis (ACA)

Some brief remarks about each of these methodological approaches are provided here as a context for the main subject of this Chapter, an author cocitation analysis.

8.2.1 Document cocitation analysis (DCA)

Working independently in the USA and the USSR respectively, Small (1973) and Marshakova (1973) simultaneously reported a new technique for exploring the natural structure of scientific specialties. This technique, document cocitation analysis (DCA), exploits the citation linkages *between individual publications* (usually scientific papers) to develop indicators of subject similarity. The principles of DCA are really quite simple and are clearly and concisely described in an article by van Raan:

> *"When a scientific paper cites two earlier papers, these latter papers are 'cocited', the strength of such a cocitation relation is determined by the number of citing papers having the above pair in their reference list. One of these cocited papers can also form a cocitation pair with a third paper. In this way, clusters of (co-)cited papers emerge, and a 'map' of the citation field can be created"* (Van Raan,1990:626 quoted in Osareh, 1996b).

Of course, when compared in this way, most document pairs return a cocitation count of zero, since no subsequent work cites them jointly. Other document pairs will return a small count, perhaps one or two. This may or may not signal relatedness; only when the number of cocitation counts rises above a given threshold can subject or other forms of relatedness be reasonably inferred.

Having established a set of pairs of documents with relatively high cocitation counts, it is then possible to visualise the relationships between them using widely-available techniques for multivariate analysis such as hierarchical clustering or multidimensional scaling. If cocitation counts are taken as measures of similarity, it follows that computer-generated maps of document cocitation will place pairs of frequently cocited documents closer together and pairs with lower counts further apart. The clusters which emerge should enable the researcher to define the boundaries of specialities and sub-specialities. If a particular cluster comprises a distinct and coherent area of scholarship, one would expect to find relatively few new papers joining that cluster as the threshold for cluster membership is reduced—at least until a certain point at which two or more clusters fuse, indicating the point at which, say, information retrieval, bibliometrics, and studies of scientific communication become incorporated into a wider intellectual movement: information science. Small (1977) notes how cluster membership thresholds may be varied in order to find the optimum solution for partitioning a set of documents. He notes however that this fine tuning is expensive and that it is no substitute for expert judgement.

Garfield (1979) presents a very useful technical overview of document cocitation methods, based on his extensive use of the technique at the *Institute for Scientific Information (ISI)*. Even for relatively small-scale studies, however, it is clear that document cocitation analysis is very highly intensive of computer and other resources and is not a methodology to be entered into lightly.

8.2.2 Journal-by-journal citation analysis

In 1989, *ISI* introduced a new field, *CW* or cited work, into its citation databases. For the first time, this made it feasible to collect numbers of cocitations *between*

journal titles rather than between individual papers as in the case of document cocitation analysis. One application of journal-by-journal citation analysis has been to determine the quality, usefulness and impact of journals (Osareh, 1996b). Rice and co-workers suggest that journal cocitations may be used to identify "forms of scientific social structure and the differential influence of different sources and disciplines of prior research" (Rice and others, 1989:258). This suggestion was taken up by Besselaar & Leydesdorff (1996) who analysed the cocitation networks between journals in robotics, electrical engineering and information science to explore the development of new paradigms in artificial intelligence research.

8.2.3 Country-by-country citation analysis

Cocitation analysis has also been used to inform 'big' science policy, for instance by providing bibliometric indicators of the output (e.g. author productivity) and the impact (e.g. citedness) of research carried out in different countries (Price, 1969; Spiegel-Rosing, 1972; Price & Gursey, 1975; Kovach, 1978; Irvine & Martin, 1989; Lancaster, 1991). The limitations of these gross bibliometric indicators are perhaps obvious. The time lag between submitting an article, then having it published and it subsequently being cited may be quite considerable, opening up the criticism that these are 'trailing' indicators of little relevance to the immediate needs of policy-makers. Other criticisms of these kinds of indicators focus on the need for a much finer level of resolution, so that the more dynamic aspects of scientific and technical research can be mapped (Irvine & Martin, 1989).

8.2.4 Author cocitation analysis (ACA)

Author cocitation analysis (ACA) was pioneered by White (1981) and by White & Griffith (1981, 1982) in a series of studies which mapped information science authors. Author cocitation studies are based on the frequency with which pairs of authors are subsequently cited by a third author. The assumption is that the more frequently two authors are cited together, and the more similar their

profile of cocitations with other authors in the set, then the closer the cognitive and / or social relationship between them.

By examining the distribution of author cocitation data within the two- or three-dimensional `intellectual space' of a mapped display, various aspects of structure can be described. Clusters of authors can be identified who share common research specialisms, schools of thought, shared intellectual styles, or who are bound together by temporal or geographic considerations. By using factor analysis it is further possible to demonstrate the breadth or concentration of an author's contribution and to identify authors who are central or peripheral to a field or speciality. Potentially, then, the technique is a powerful methodology for mapping areas of scholarly activity.

McCain (1990) argues that the techniques and assumptions of author cocitation analysis are closely related to document cocitation studies and the mapping techniques associated with the work of Small, Griffith and co-workers. Author cocitation analysis is best thought of as the analysis of highly cocited pairs of *oeuvres*, rather than pairs of individual documents. Although ACA studies thus operate at a higher level of abstraction than DCA studies, there are significant practical benefits in choosing author cocitation as a research methodology. The most notable advantage is the ease with which it is possible to obtain cocitation counts online. These may be obtained simply by combining sets of cited authors (using AND logic). Indeed, the ease with which this data may be collected is such that it has led White & Griffith to argue that ACA represents "a technical advance that permits the mapping to order of any small field or specialty an investigator cares to name" (White & Griffith, 1981:164).

8.3 Mapping fields of scholarship

Implicit in terms such as `areas' or `fields of study' there is a spatial connotation. Vannevar Bush recognised this when he coined the metaphor of the scientific `frontier' (Bush, 1945). DeSolla Price (1966), a pioneer in the application of bibliometric techniques for scientometric purposes, showed how the topography of science could be represented in the form of two-dimensional maps, revealing networks of scientific papers linked by citation. More recently, Small & Garfield (1986) describe how the relationships between scientific disciplines can be mapped at a micro-level in terms of the explicit citations linking documents and authors, and at a macro-level in terms of common methodologies and concepts linking communities of scholars. Their studies, like those of Tijssen and co-workers, draw on a range of non-bibliometric sources of information:

> *"The last decades have seen an increasing utilisation of graphical representations of aspects of science. These `maps' are mainly used to depict the underlying relational structures of publications, or publishing entities, within the science and technology system. To this end, the maps generally draw on only one source of bibliometric (i.e. literature-based quantitative) data: mostly citations or keywords. These information items will necessarily only describe one facet of the (intellectual and/or social) structure of science. We argue that in order to obtain a more complete description of the common underlying structure one requires the incorporation of more sources of data"* (Tijssen and others, 1990:224).

In practice, similarities between authors or published articles cannot be fully represented by citation relationships alone. Perhaps a more important consideration is the degree of similarity in content. Intuitively, it seems reasonable that by incorporating content-based indicators into a literature-based mapping exercise, a more complete description of author-author or article-article similarities should be possible.

Tijssen and others (1990:226) conclude that "...in addition to relational data such as citations, one may want to include information of a more structural nature,

such as expert opinions, or contents profiles of journals". Bellardo also notes that the validity of cocitation analysis is imperfect, but that it is "strong when used in conjunction with other measures and indicators" (Bellardo, 1980:231). Among the techniques which have been used to explore the external validity of author cocitation data are questionnaires, intellectual histories based on review articles, and short histories of particular events in the field. These approaches attempt to get at the same information from a different perspective in order to confirm the results. The proper conclusion seems to be that every method has its weaknesses and biases, but at least these are different.

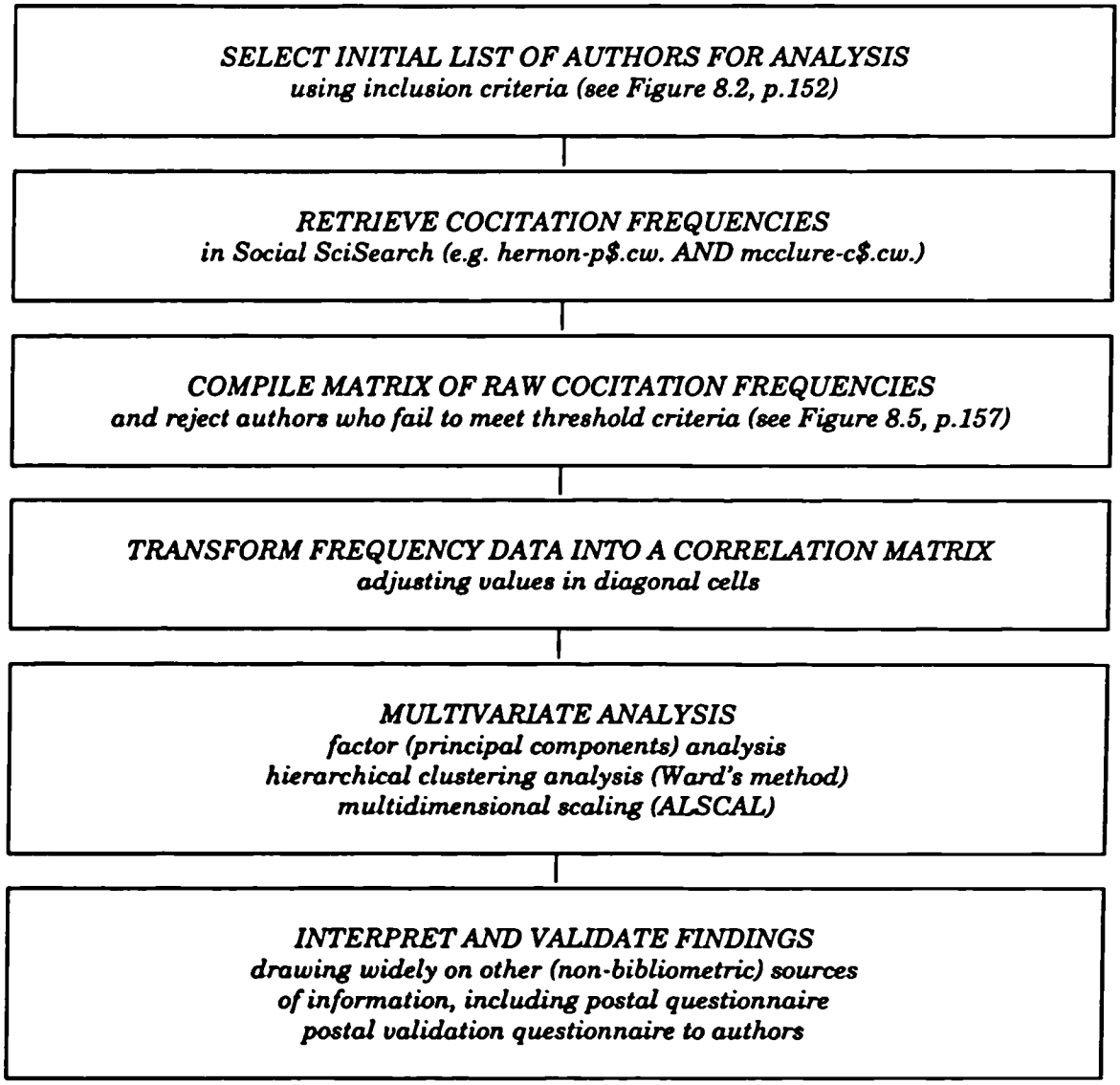
For these reasons, the interpretation of the experimental results presented later in this Chapter draws heavily on expert opinion and external non-bibliometric sources of information.

8.4 Research design and methodology

The research design for this study is shown schematically in Figure 8.1 overleaf. It determines the structure for the rest of this Chapter.

The research design closely replicates the methodology of White & Griffith (1981, 1982). As well as being the pioneering work in this area, the research reported by White & Griffith in 1981 was felt to be particularly relevant to this study—it covered 39 key authors from the field of information science. Some account has also been taken of more recent authors, notably Bayer, Smart & McLaughlin (1990); McCain (1990); and Kärki (1996).

Figure 8.1: Research design



8.4.1 Selection of the initial author set

The first stage in the research was to identify a list of candidate authors whose cited references could be retrieved from *Social SciSearch*. This represents possibly the most critical and certainly the most contentious stage in setting up an author cocitation study. It is essential to establish a diversified list of authors so that the full richness of the scholarly landscape can be captured. If the authors chosen do not reflect the full range of variability in subject

specialisations, research perspectives and institutional affiliations then, of course, these aspects of structure cannot be demonstrated.

Previous author cocitation studies have employed a variety of approaches in generating an initial list of candidate authors. In each case, the guiding philosophy has been to use as neutral and objective a set of inclusion criteria as possible. In their 1981 study of information scientists, White & Griffith drew their authors from a collection of seminal articles, *Key Papers in Information Science* (Griffith, 1980). Other workers have identified prominent authors by examining such primary and secondary sources as:

- monographs
- reviews of books and monographs
- review articles
- authors of encyclopaedia entries
- lists of academic prizewinners
- members of prestigious academic or professional bodies
- lists of conference speakers
- editorial board members (or referees) of leading journals
- entries in *Who's Who* type publications
- holders of large research grants and contracts

Often, several documentary sources are used in combination. Other, more experimental, approaches have been adopted. Cottrill, Rogers & Mills (1989), for example, used a snowball citation search method to identify highly cited *oeuvres*

in the technology transfer research literature, while Bayer, Smart & McLaughlin (1990) used a questionnaire survey method to establish a list of highly influential scholars.

The use of secondary data sources to identify candidates has the advantage that it is reasonably neutral and objective. Personal judgement is still an important feature of many author cocitation studies, however, both in the selection of appropriate secondary sources and in any subsequent decisions to add to or delete names from the list. Thus, White & Griffith's 1981 study of information scientists used not only the 22 authors drawn from the *Key Papers* volume, but a further 17 whom they judged to be `major contributors to the field', including such early pioneers as Shannon, Zipf and Luhn. Unfortunately, no predetermined list of major contributors in information policy could be identified for the purposes of this study. Instead, the candidate authors for this study were derived from the experimental test collection using the criteria set out below in Figure 8.2.

Figure 8.2: Initial author selection criteria

All authors must be represented in the document test collection

AND EITHER

Be located in Bradford Zone 1 of author productivity

OR

Satisfy at least two of the following criteria:

- *have published and / or been cited over a ten-year period*
- *have attracted at least 40 citations in Social SciSearch*
- *be judged to have made a significant contribution to the field*

In all 34 initial authors were recovered in this way: each of the 22 authors who appeared in Bradford zone 1 of author productivity in the document test collection, plus a further 12 selected on the basis of personal judgement and knowledge of the information policy literature (see Figure 8.3 overleaf). Any sins

of commission and omission, clearly a potential issue in studies of this nature, therefore lie entirely with the author of this thesis.

Figure 8.3: Initial list of candidate authors

*Anthony, L.J.**	*Eisenbeis, K.**	*Kirtley, J.E.**	*Oppenheim, C.*
Bearman, T.C.	*Feinberg, L.E.**	*Love, J.P.**	*Regan, P.M.**
Bortnick, J.	*Flaherty, D.H.*	*Mansell, R.E.**	*Relyea, H.C.**
*Braman, S.**	*Gillham, V.*	*Martyn, J.*	*Rosenberg, V*
Burger, R H.	*Griffith, J.C.**	*McClure, C.R.**	*Sauvant, K.P.**
*Case, D.O.**	*Hernon, P.**	*Moore, N.**	*Shill, H.B.*.*
Chartrand, R.L.	*Hill, M.W.*	*Morehead, J.*	*Sprehe, J.T.**
Cronin, B.	*Irwin, M.R.**	*Morton, B.**	
*Durrance, J.C.**	*Katz, J.E.**	*O'Reilly, J.T.**	

*Author in Bradford zone 1 of author productivity

8.4.2 Online data collection

Cocitation frequencies for each pair of authors in the initial list were collected by searching *Social SciSearch* on *DataStar*. Each of the 34 authors' names was searched in turn to create postings reflecting all the publications citing that author over the period 1972 - January 1997, regardless of format (articles, conference papers, monographs, book reviews).

So, in the case of Peter Hernon, for example, the search statement was:

HERNON-P$.CW.

The number of works cociting Hernon and a second author, Charles McClure, was then simply obtained from the intersection of the individual author sets:

HERNON-P$.CW. AND MCCLURE-C$.CW.

The total number of cocitation frequencies retrieved in this way was $n(n-1)/2$, where n is the number of authors. In this case the total number of unique author pairs was 34(33)/2 = 561.

Examples of some recent information policy works jointly citing Hernon and McClure may be inspected in Figure 8.4. Note that in line with previous ACA

studies, self-citations are included. This is arguably a less objectionable practice in cocitation as opposed to citation studies; the principle being that if an author consistently cites him- or herself with another author, a positive cognitive or social relationship may be inferred.

Figure 8.4: Examples of works cociting Hernon and McClure

Bergeron, P (1996)
Information resources management,
Annual Review of Information Science & Technology 31 263-300.

McClure, CR (1996)
Libraries and Federal information policy
Journal of Academic Librarianship 22(3) 214-218.

Kajberg, L & Kristiansson, M (1996)
An overview of the field of information policy
International Forum on Information & Documentation 21(1) 5-9.

Braman, S (1995)
Policy for the Net and the Internet
Annual Review of Information Science & Technology 30 5-75.

Rawan, AR & Cox, J (1995)
Government publications: integration and training
Journal of Government Information 22(3) 253-266.

Moen, WE (1994)
Information technology standards and Federal information policy
Government Information Quarterly 11(4) 357-371.

Smith, TD (1994)
Measuring the effect of US OMB Circular A-130
Journal of Government Information 21(5) 391-402.

The limitations of citation searching on *ISI* databases are well known and are discussed in more detail in the concluding sections of this Chapter. Two important limitations need to be discussed immediately, however. The first is that it is only possible to collect citation and cocitation information for sole or first-named authors. Hence, scholars who collaborate with others but who do not obtain first authorship are not represented. Garfield (1979) argues, however, that since co-author order is based on the relative importance of an individual's contribution, this is a relatively minor problem.

The second issue arises from the fact that *ISI* indexes author surnames and initials only. Consequently, for common surnames such as Hill or Griffith there is a possibility that the online cocitation retrieval procedures described above will pick up name homographs. In the search example above, it will be noted that a truncation symbol was used to generalise the request. Thus the search request *MCCLURE-C$.CW.* will retrieve any materials that cite work by C. McClure, whether or not the citing authors used his second initial ('R'). There is a danger of course that papers citing the 'wrong' C. McClure will also be retrieved. This issue appears to be less important in practice than one might imagine: White & Griffith estimated an error rate due to name homographs of less than two per cent during their 1981 study and a preliminary examination of a sample of cocitations from the present study suggested a similar error rate, although with one significant exception—the documents cociting M$Hill and J$Griffith were found to relate entirely to articles on the criminal justice system rather than information policy! These were naturally removed from the analysis.

8.4.3 Compilation of the raw data matrix

When the raw cocitation frequencies were inspected, and as might have been expected, not all pairs of authors were found to have substantial cocitation counts—indeed in some cases the counts were very small. This raises an important issue concerning the potential instability of author cocitation data. Low cocitation counts may or may not be indicative of a cognitive or other relation between two cited authors. Only in the case of relatively high cocitation counts may one suspect and then further investigate a possible direct relationship. In a highly coherent disciplinary area, and given an appropriate selection of authors, one would expect to find a reasonably large proportion of high cocitation counts and relatively few occasions where author pairs scored mostly zeros (and could not therefore be reasonably said to be fully integrated into the set as a whole). Previous researchers have used a variety of *ad hoc* threshold criteria to screen an initial author list, based on three measures: citation frequencies, connectedness and mean cocitation rate.

Cocitation frequencies

In their 1982 paper, White & Griffith rejected any author who had fewer than 45 cocitations in total (this arbitrary threshold was based on the observed distribution of cocitation frequencies in their dataset). Another researcher, Penan, restricted his final author set to the upper quintile receiving the highest number of citations and cocitations in the initial retrieval trials (Penan, 1989 cited in McCain, 1990).

Connectedness

In this thesis, the term 'connectedness' refers to the proportion of non-zero cocitation counts for each author pair. Given an initial sample of 34 authors, each author has an opportunity to be cocited with any or all of the remaining 33. Hence, an author with 33 non-zero cocitation counts would be said to show 100 per cent connectedness.

The concept of connectedness is a useful tool for screening out candidates who integrate loosely, sometimes barely at all, with the remaining body of authors. White & Griffith used a connectedness threshold value of 33 per cent in their 1982 study.

Mean cocitation rate

A further screening procedure used in some author cocitation studies is the mean cocitation rate—the arithmetic mean of all the cocitation frequencies in the matrix. As a rule of thumb, White & Griffith (1982) recommend a minimum mean cocitation rate of nine (for ten years of *Social SciSearch* data). McCain (1990), however, reports experimenting with mean cocitation rates as low as four (across five years of *Social SciSearch* data) and found the results to be satisfactory in terms of stability, face validity and interpretability.

On the basis of these considerations, a series of threshold criteria was developed for the present study (see Figure 8.5):

Figure 8.5: ACA threshold criteria used in the present study

Each author must have a total of at least 40 cocitation counts

Each author must be cocited with at least 1/3 of the other authors

The mean cocitation rate for the whole set must be > 4

After a number of iterations, the initial list of 34 authors was reduced to a final set of 21 authors which was fully compliant with the above criteria. This is a rather small residual set, at least by comparison with other published studies, a point which will be discussed later in this Chapter.

The final list of authors is shown in Table 8.2. Most of the authors exhibit a high degree of connectedness, although it is noted that the mean cocitation rate (5.0) is low, at least in comparison with published studies from other literatures.

Table 8.2: Authors included in the final study

Author	Number of cocitations	% Connectedness with other authors
Hernon	401	100
Martyn	121	100
Bearman	69	95
Cronin	187	90
Chartrand	84	90
Bortnick	54	90
Burger	48	90
McClure	326	85
Morehead	118	80
Relyea	75	80
Case	56	80
Morton	115	75
Moore	86	75
Rosenberg	56	75
Hill	50	75
Katz	41	75
Oppenheim	66	70
Anthony	49	70
Flaherty	43	55
Sprehe	88	50
Durrance	69	50

In the final stage of the initial data preparation, the cocitation counts for each pair of authors were arranged in the form of a symmetrical matrix with

identically ordered authors' names on the rows and columns. This is illustrated in Table 8.3 below for the sake of clarity.

Table 8.3: Author cocitation frequencies matrix

	1																			
1	*10*	2																		
2	2	*12*	3																	
3	2	4	*8*	4																
4	1	3	2	*7*	5															
5	0	1	1	1	*12*	6														
6	4	2	2	3	4	*15*	7													
7	4	11	1	2	5	3	*37*	8												
8	0	1	0	2	0	0	5	*19*	9											
9	0	0	2	0	1	2	0	0	*10*	10										
10	3	4	8	6	10	14	12	21	3	*85*	11									
11	0	1	1	2	0	3	3	0	1	9	*3*	*12*								
12	1	1	2	2	1	0	1	1	13	2	1	*9*	13							
13	2	2	4	4	8	10	20	12	3	110	1	1	*74*	14						
14	10	9	1	2	3	2	29	2	0	3	3	1	3	*25*	15					
15	0	4	1	1	1	2	24	3	1	5	5	2	8	7	*20*	16				
16	0	2	3	2	1	2	5	2	3	29	2	0	17	3	1	*28*	17			
17	1	1	3	2	2	4	3	0	0	31	1	0	18	0	0	9	*29*	18		
18	1	1	0	3	3	2	17	1	0	2	4	0	3	10	1	0	1	*16*	19	
19	2	3	5	2	1	1	2	0	0	15	1	2	7	1	0	4	3	0	*15*	20
20	5	4	2	1	1	3	3	0	2	6	0	1	5	7	0	0	2	1	4	*9* 21
21	1	1	2	0	0	6	0	0	2	23	0	0	14	0	0	5	5	0	7	0 *22*

Key to authors

1	Anthony	8	Durrance	15	Moore
2	Bearman	9	Flaherty	16	Morehead
3	Bortnick	10	Hernon	17	Morton
4	Burger	11	Hill	18	Oppenheim
5	Case	12	Katz	19	Relyea
6	Chartrand	13	McClure	20	Rosenberg
7	Cronin	14	Martyn	21	Sprehe

Following the example of White & Griffith (1981, 1982) and many other workers, this raw data matrix was subsequently transformed into a similarity matrix of Pearson product-moment correlation coefficients. The rationale for this step is explained in the next section. However, in order to be able to accomplish this step successfully in *SPSS*, an immediate problem which needed to be tackled was which values should be placed in the empty diagonal axis of the matrix? These cells represent the intersection of a particular author with him- or herself (e.g. Oppenheim-Oppenheim) and it is not clear at all what values, if any, should be placed here.

White & Griffith (1981) report some difficulty in resolving this issue. In one trial, they filled the diagonal cells with the total number of cocitations for each author, but found that the values were often disproportionately large, sometimes an order of magnitude greater than any cocitation count in the off-diagonal cells. After some deliberation, White & Griffith arrived at a formula which approximates, for each author, the value of what, hypothetically, would be the *next highest score* in the distribution. This was determined by adding together the three highest cocitation counts for each author and dividing by two.

In an alternative approach, McCain claimed that there was "little difference, in mapping, clustering, and factor analysis between scaling the diagonal values (à la White & Griffith) and treating them as missing data" (McCain, 1990:435). Bayer, Smart & McLaughlin (1990) tackled the problem by transforming their raw cocitation frequencies into a rank order dissimilarity matrix, where each diagonal cell was given a rank of one, indicating that each scholar was `most like' or closest to him- or herself.

Given the difficulty of finding a non-arbitrary value for the diagonal cells, the present study simply adopts the White & Griffith convention (these scaled values are indicated in bold italics in Table 8.3).

8.4.4 Generation of correlation profiles

As already noted, the final data preparation stage involved transforming the raw cocitation frequencies into a matrix of Pearson product-moment correlation coefficients, *r*. This transformation offers two major advantages. Firstly, it provides a means of normalising the data. Inspection of Table 8.3 reveals that the cocitation frequencies range over two orders of magnitude (minimum = 0, maximum = 110). While the data in Table 8.3 offer a very direct way of measuring the absolute degree to which two authors are cocited, a practical difficulty emerges when projecting very large and very small frequencies onto a two-dimensional space using ALSCAL procedures (see Section 8.5.3, p.168). When represented in this way, extreme values tend to force a bunching effect on the resulting maps which may make them almost impossible to read and

interpret (see Kerlinger, 1973). The approach taken in this study has been to use both the transformed and the original raw data matrices for the purposes of guiding the final interpretation.

The second major advantage of creating a matrix of Pearson correlation coefficients is that it provides richer information than cocitation frequencies. For any given pair of authors, the Pearson coefficient functions as a measure not just of how often that particular pair are cocited (i.e. the cocitation count) *but how similar their cocitation profiles are, taken across the whole author set.* This may be best illustrated by example. Table 8.4 shows the cocitation counts for two author pairs: Hernon-McClure and Hernon-Martyn. It has been abbreviated for ease of assimilation.

Table 8.4: Partial cocitation counts for two author pairs

	Hernon	McClure		Hernon	Martyn
Anthony	3	2	Anthony	3	10
Bearman	4	2	Bearman	4	9
Bortnick	8	4	Bortnick	8	1
Burger	6	4	Burger	6	2
Case	10	8	Case	10	3
Chartrand	14	10	Chartrand	14	2
Cronin	12	20	Cronin	12	29
Durrance	21	12	Durrance	21	2
Flaherty	3	3	Flaherty	3	0
Hill	9	1	Hill	9	3
Katz	2	1	Katz	2	3
...	...	...	...	...	...
Rosenberg	6	5	Rosenberg	6	7
Sprehe	23	14	Sprehe	23	0
		$r = 0.92$			$r = -0.21$

Hernon and McClure show a high positive correlation (r = 0.92) not primarily because they are highly cocited with each other but because they tend to be cocited frequently or infrequently *by the same authors.* In the case of Hernon-Martyn, the reverse is true and they tend to be cited frequently or infrequently by third authors. Hernon's 'cocitation profile' with Martyn is very different from that with McClure and is in fact negative (r = -0.21). Within the conventions of author cocitation analysis, these findings suggest that Peter Hernon is much 'closer' to Charles McClure in terms of social and / or cognitive distance than he is to John Martyn.

The Pearson correlation matrix is depicted in Table 8.5, again as a lower diagonal half-matrix. For ease of interpretation, decimals are omitted and only values of r >= 0.4 that are associated with a high level of significance, $p < 0.05$, are shown.

Table 8.5: Similarity matrix of Pearson correlation coefficients, *r*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	100																				
2	.	100																			
3	.	.	100																		
4	.	.	.	100																	
5	.	.	.	.	100																
6	.	.	.	.	56	100															
7	.	65	.	.	.	.	100														
8	.	.	.	46	.	.	.	100													
9	.	.	.	.	.	.	.	.	100												
10	.	.	47	44	56	63	.	68	.	100											
11	.	.	.	.	.	.	.	.	.	.	100										
12	.	.	.	.	.	.	.	.	86	.	.	100									
13	61	75	.	.	.	.	84	.	.	.	.	.	100								
14	.	.	54	52	64	69	.	77	.	92	.	.	.	100							
15	.	54	.	.	.	.	84	.	.	.	.	.	64	.	100						
16	.	.	52	.	.	45	.	53	.	76	.	.	.	78	.	100					
17	.	.	54	.	46	57	.	47	.	77	.	.	.	80	.	.	100				
18	.	.	.	.	.	.	74	.	.	.	.	.	78	.	49	.	.	100			
19	.	.	76	.	.	.	.	.	.	57	.	.	.	63	.	56	54	.	100		
20	74	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	
21	.	.	52	.	.	67	.	.	.	73	.	.	.	75	.	64	65	.	73	.	100

Key to authors

1	Anthony	8	Durrance	15	Moore
2	Bearman	9	Flaherty	16	Morehead
3	Bortnick	10	Hernon	17	Morton
4	Burger	11	Hill	18	Oppenheim
5	Case	12	Katz	19	Relyea
6	Chartrand	13	McClure	20	Rosenberg
7	Cronin	14	Martyn	21	Sprehe

The data in Table 8.5 may be regarded as measures of similarity between pairs of authors in terms of the collective perceptions of those cociting them. This matrix constituted the *SPSS* input file for the next stage of the study: multivariate data analysis.

8.4.5 Postal questionnaire

In parallel with these experiments, a simple postal questionnaire was developed (see Appendix F) and sent to each of the 21 authors featuring in the final study. The questionnaire was designed to validate certain aspects of the author cocitation study; the results are discussed in Section 8.7.

8.5 Multivariate analysis

Three multivariate tools were used to explore the data in the correlation matrix: factor (principal components) analysis; cluster analysis; and multidimensional scaling (MDS). These are complementary techniques, each offering different insights into the structure of the data. In each case, the analyses were was carried out using standard facilities provided in *SPSS* Version 6.1.

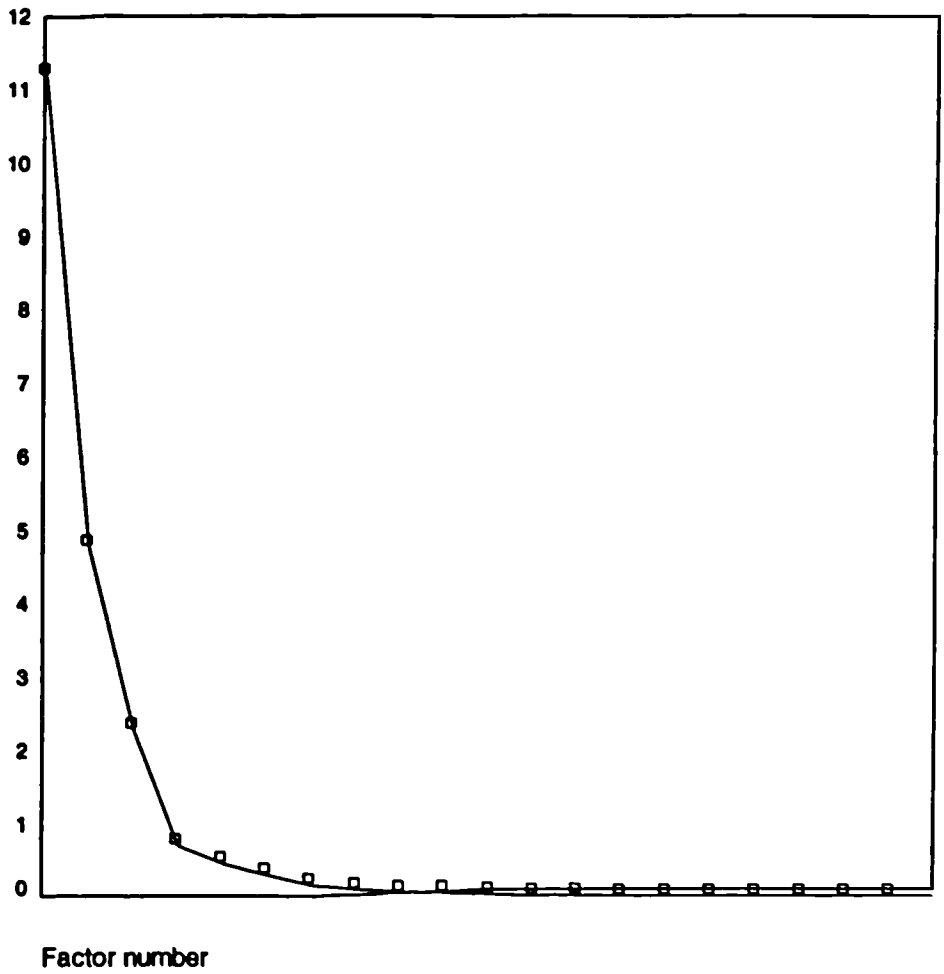
8.5.1 Factor (principal components) analysis

Factor analysis attempts to clarify and explain the interrelationships between the observed variables in a data set by creating a much smaller set of derived variables. These derived variables or *factors* can help to explain underlying dimensions in the observed data. In author cocitation studies, factor analysis is typically used to reveal the underlying 'hidden' subject matter as perceived by the population of citing authors. In this study, the correlation matrix was factor analysed[54] and rotated using the principal components analysis (PCA) routines in *SPSS*.

[54] A measure of the appropriateness of carrying out a factor analysis on a correlation matrix is provided by the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. Kaiser (1974) characterises measures in the .90s as *marvellous*, in the .80s as *meritorious*, in the .70s as *middling*, in the .60's as *mediocre*, in the .50's as *miserable*, and below 0.5 as *unacceptable*. The KMO value obtained in this case was 0.77, suggesting that there is some justification for proceeding with a factor analysis.

One very useful feature of *SPSS* is that it automatically generates a graphic, a factor scree plot, which enables the researcher to arrive quickly at the most parsimonious solution in terms of the number of factors which need to be considered. The plot is said to resemble a steep cliff with a shallower sloping bed of rubble, or scree, at its foot. Figure 8.6 displays the Eigenvalues (y axis) associated with each of a maximum of 20 factors (x axis) arranged in decreasing order. The plot moves sharply from 'cliff' to 'scree' between the fourth and the fifth factors. In fact, the four largest extracted factors account for 92.1 per cent of the variance in the correlation matrix and so a four-factor solution was judged to be the most parsimonious.

Figure 8.6: Factor scree plot



The appropriateness of a four-factor solution is given added weight by the fact that the authors exhibit a high degree of communality under these conditions (minimum value, 0.77), further indicating that no significant latent variables have been missed.

The four 'hidden' factors revealed by principal components analysis are presented Table 8.6 in order of the total variance accounted for, Factor 1 being the highest. Within each column, the authors loading most heavily on each factor are listed (above an arbitrary threshold of 0.35). The findings are taken from the 'structure matrix' of the *SPSS* output.

Table 8.6: Oblique factor analysis: author factor loadings at 0.35 or higher

Factor 1: ***"Government Information"***		**Factor 2:** ***"Scientific & Technical Information"***		**Factor 3:** ***"Social Implications of ICTs"***		**Factor 4:** ***"Information Infrastructure & Regulation"***	
McClure	.97	Rosenberg	.89	Katz	.62	$Oppenheim_2$	.95
Hernon	.96	$Anthony_1$	.85	Flaherty	.53	$Cronin_2$	.94
Morton	.91	$Bearman_1$	.68			Moore	.93
Durrance	.89	$Martyn_1$	.54			$Martyn_2$	.91
Morehead	.89	$Cronin_1$	.45			$Bearman_2$	.75
Case	.88	$Oppenheim_1$	.37			$Anthony_2$	.43
Chartrand	.88						
Sprehe	.85						
Relyea	.72						
Burger	.70						
Bortnick	.60						

Note: Some authors load on more than one factor, indicated by subscripts.

Each factor has been given a label which provisionally summarises the subject perceptions of the citing authors. These were inferred by examining the titles of the relevant citing articles. Authors loading on Factor 1 are identified as

specialists in government information resources, all of them American. These authors have written extensively on the policy issues associated with the management, control and dissemination of (mainly) Federal information resources. Factor 2 picks out a group of authors with common interests in scientific and technical information policy. The citing papers associated with this factor appear to be considerably older than those associated with the other three factors.

Factor 4 exhibits considerable overlap with Factor 2, with five authors common to both groups. However, when the citing references are examined, it is clear that the two factors relate to different topics: Factor 4 is associated with articles that relate to the broad policy issues associated with the information infrastructure and its regulation (including such topics as information markets, public-private synergy, trade in information services, and information law). The citing articles tend to be much more recent than those associated with Factor 2. The links between these two factors show how important it is to ground an interpretation firmly in the information content of the citing articles, not just upon one's subjective (and possibly incomplete) impressions of the interests of the *oeuvres* concerned.

Factor 3 relates to two authors (David Flaherty and James Katz) who are media and communications studies specialists with a critical interest in the social implications of information technologies, especially privacy concerns. With the exception of Tony Anthony and Victor Rosenberg, all the remaining LIS authors loaded negatively on this factor.

One advantage of factor analysis over graphical multivariate display techniques is that it is possible to show authors who load heavily on more than one factor—this is clearly impossible when the same data is represented as a map or a dendrogram. Authors loading heavily on more than one factor are indicated in Table 8.6 using the convention of a subscript. It can be seen, for example, that Toni Carbo Bearman loads both on Factor 2 (Scientific and Technical Information policy) and on Factor 4 (Information Infrastructure and Regulation).

Oblique factor analysis therefore can shed light not only on an author's perceived contribution to a specific area but also on the *breadth* of their contribution.

Factor analysis also provides further clues about the structure and coherence of a field or domain. For example, it is possible to show how dependent or independent the factor groupings are relative to one another by constructing a factor intercorrelation matrix (see Table 8.7):

Table 8.7: Factor intercorrelation matrix

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	1.00			
Factor 2	0.04	1.00		
Factor 3	-0.43	-0.04	1.00	
Factor 4	-0.41	0.22	-0.05	1.00

McCain offers the rule of thumb that "in highly coherent fields, certain factors may have intercorrelations of 0.3 or above, pointing to links between research specialities or other constructs" (McCain, 1990). The data in Table 8.7 suggest that Factors 1 and 3 are uncorrelated (and therefore `independent') and that Factors 2 and 4 are only weakly co-dependent. On the basis of these indicators alone, one might well conclude that information policy does not represent a coherent field (or at least, if it does, that the subject relationships are not yet paradigmatic).

8.5.2 Hierarchical cluster analysis

Cluster analysis is a useful tool for examining the presence of natural structure in the correlation matrix by grouping author profiles according to their similarity. Many published author cocitation studies have employed a hierarchical agglomerative (`bottom up') approach to clustering, using single linkage, complete linkage, average linkage, or, more usually, Ward's method. One problem with hierarchical clustering is that there are no generally agreed

stopping rules to alert the researcher to the `best' set of clusters to report—i.e. the number of clusters that most closely corresponds with the underlying natural structure in the data (see, for example, Ling & Killough, 1976; Dubes & Jain, 1979, Bailey & Dubes, 1982; Aldenderfer & Blashfield, 1983; Rowlands, 1983). This is not necessarily a significant problem in author cocitation studies. Here, the main goal is to inform a more general discussion of the relationships between authors—the `true' number of clusters in the matrix is perhaps of less interest.

Typically, ACA researchers choose a single cluster level for detailed analysis and then refer `down' to subclusters or `up' to macroclusters where this is useful. Figure 8.8 on the next page was prepared in this spirit, showing cluster group membership at various thresholds. A dendrogram resulting from clustering[55] the variables in the correlation matrix is also displayed as Figure 8.7.

[55] The hierarchical cluster analysis presented here followed Ward's method, using a simple Euclidean measure of distance.

Figure 8.7: Dendrogram (Ward's method)

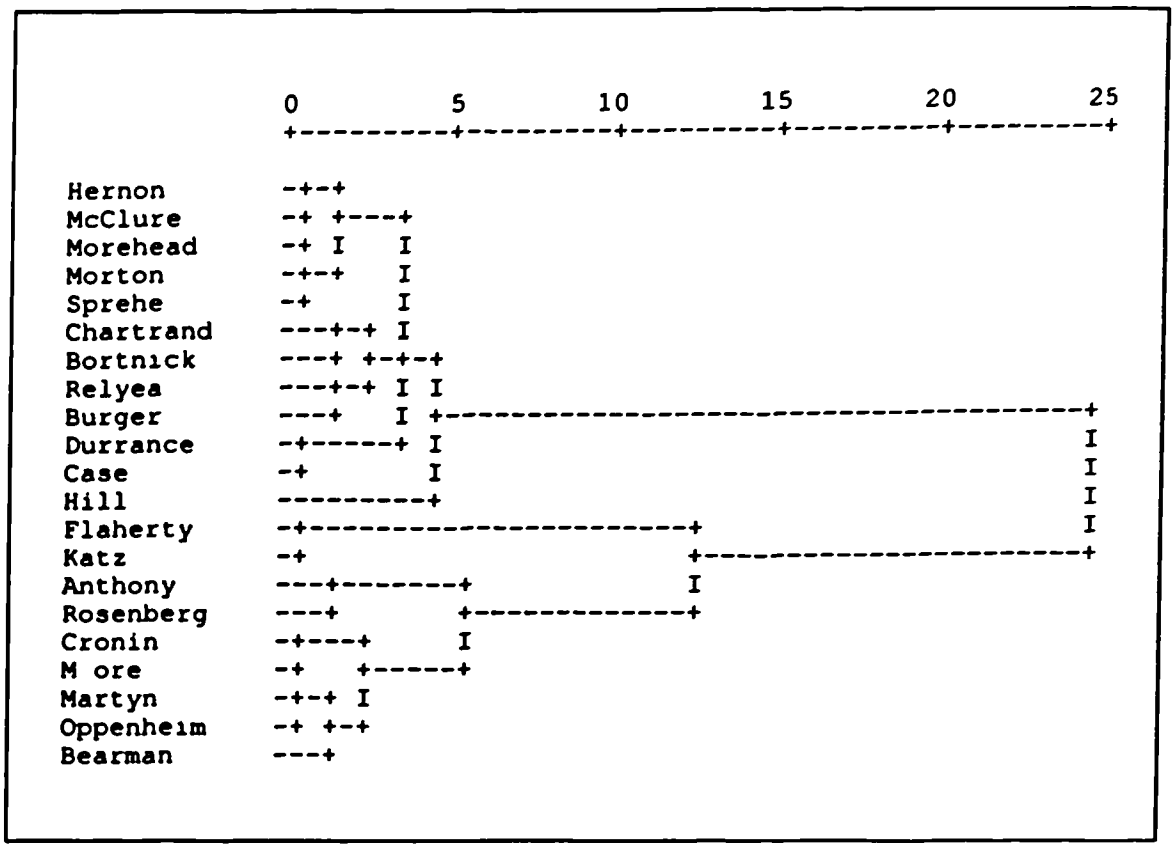


Figure 8.8: Cluster membership (Ward's method)

Three Cluster Groups	Four Cluster Groups	Five Cluster Groups	Six Cluster Groups
Cluster 1: Anthony, Bearman, Cronin, Martyn, Moore, Oppenheim, Rosenberg	***Cluster 1:*** Anthony, Rosenberg	***Cluster 1:*** Anthony, Rosenberg	***Cluster 1:*** Anthony, Rosenberg
Cluster 2: Bortnick, Burger, Case, Chartrand, Durrance, Hernon, Hill, McClure, Morehead, Morton, Relyea, Sprehe	***Cluster 2:*** Bearman, Cronin, Martyn, Oppenheim, Moore	***Cluster 2:*** Bearman, Cronin, Martyn, Moore, Oppenheim	***Cluster 2:*** Bearman, Cronin, Martyn, Moore, Oppenheim
Cluster 3: Flaherty, Katz	***Cluster 3:*** Bortnick, Burger, Case, Chartrand, Durrance, Hernon, Hill, McClure, Morehead, Morton, Relyea, Sprehe	***Cluster 3:*** Bortnick, Hernon, McClure, Morehead, Morton, Relyea, Sprehe	***Cluster 3:*** Bortnick, Relyea
	Cluster 4: Flaherty, Katz	***Cluster 4:*** Burger, Case, Chartrand, Durrance, Hill	***Cluster 4:*** Burger, Case, Chartrand, Durrance, Hill
		Cluster 5: Flaherty, Katz	***Cluster 5:*** Flaherty, Katz
			Cluster 6: Hernon, McClure, Morehead, Morton, Sprehe

It is interesting to compare the results of a four-cluster solution with those of the factor analysis reported earlier. A very close mapping is possible between

membership of Cluster 3 and the authors who loaded most heavily on Factor 1 (Government Information). Similarly, there is a close affinity between Factor 3 (Social Implications of IT) and Cluster 4.

Factors 2 (Scientific and Technical Information Policy) and 4 (Information Infrastructure and Regulation) seem to relate closely to Clusters 1 and 2, although the cluster analysis partitions Anthony and Rosenberg (the two highest loading authors on Scientific and Technical Information Policy) from the others at an early stage.

The close affinity between the four-factor solution and the four-cluster solution is encouraging. Cluster 3 is however quite large, and it may be possible to derive a final interpretation by splitting Cluster 3 into two or more subclusters (this is explored in Section 8.6 where some external, non-bibliometric sources of evidence are introduced). For the moment, though, it should be noted that the finest level of resolution at which there are no singleton authors is six clusters.

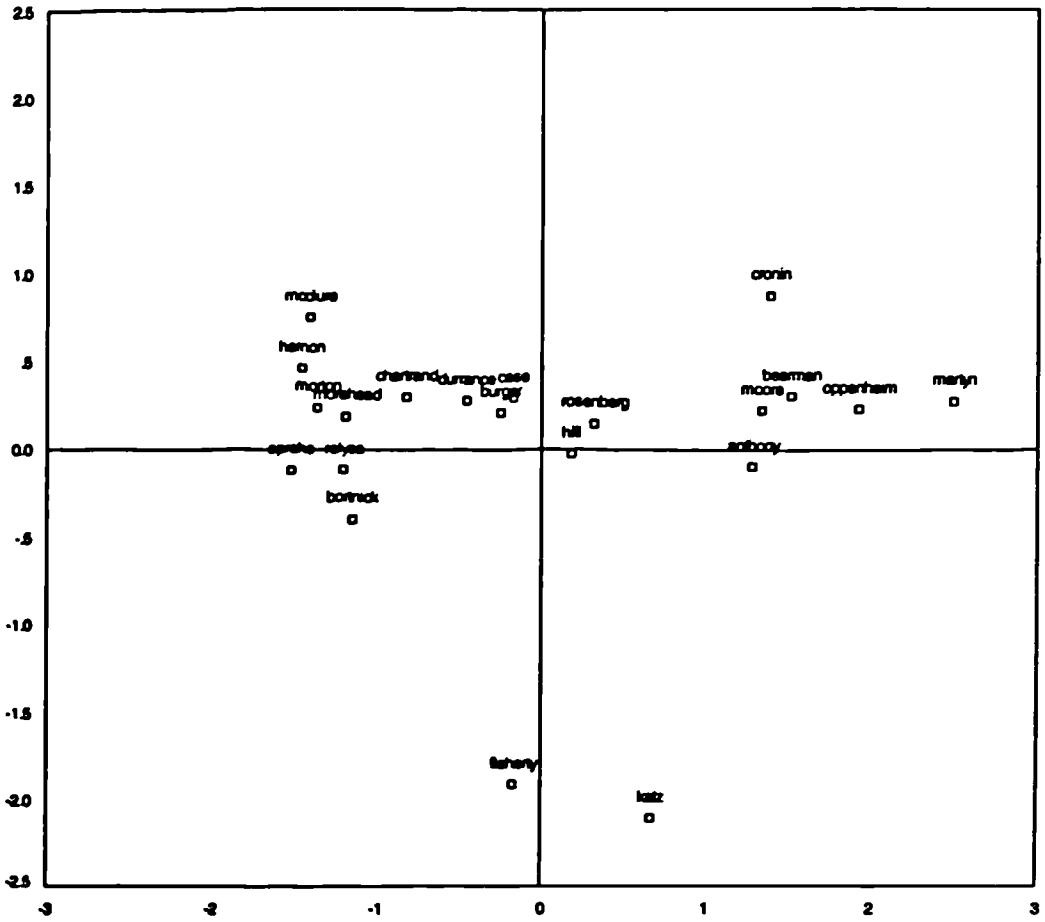
8.5.3 Multidimensional scaling

Multidimensional scaling refers to a set of techniques which project complex data onto a two- or three-dimensional space so that the relationships between variables (in this case authors) can be visualised graphically.

As before, the data input was the Pearson correlations matrix. A two-dimensional visual mapping was generated (see Figure 8.9 overleaf) using the ALSCAL routines[56] within *SPSS* Version 6.1.

[56] The ALSCAL mapping was produced using a nonmetric approach, an ordinal level of measurement and a Euclidean distance model for plotting points, following the example of McCain (1990). 'Goodness of fit' statistics were excellent: the scatterplot of distances vs disparities was strongly linear and the final two-dimensional model exhibited low stress (Kruskal stress formula $I = 0.23$) and explained a high proportion of variance ($R^2 = 0.92$).

Figure 8.9: ALSCAL mapping of information policy authors



The application of multidimensional scaling techniques in author cocitation studies is very widespread; indeed it is difficult to imagine how such studies could have been carried out before these techniques became available.

In developing an interpretation based on Figure 8.9, a number of points need first to be considered. The underlying assumption is that pairs of authors with similar cocitation profiles will tend to cluster together on the computer map. Similarly, pairs of authors with low or negative cocitation profiles will tend to be placed relatively far apart. Figure 8.9 shows, for instance, that Peter Hernon and Charles McClure, who are heavily cocited (110 times) and who share very similar correlation profiles (r = 0.92) are placed close to one another. On the other hand, Peter Hernon and John Martyn (jointly cited only 3 times, r = -0.21) are displayed at opposite extremes of the x axis.

White & Griffith (1980) argue that plots generated from highly patterned cocitation data may be used to draw a number of inferences about the intellectual and social structure of a research specialty. These include the identification of:

- coherent author groups, akin to `schools'
- the centrality or peripherality of individual authors, both with respect to their immediate author group and to the field as a whole
- locations of author groups with respect to each other
- the positions of authors and schools with respect to the map's axes

It should be noted that in multidimensional scaling procedures, the origin and axes of the resulting plots are set automatically and in an apparently `arbitrary' fashion by the software. In many published interpretations, however, the author located closest to the origin is taken to be the most `central' or representative of the field as a whole. Given the ambiguity associated with the plot axes, one of the first objectives in any interpretation of author cocitation data is therefore to make some sense of what meaning these axes convey.

8.6 Interpretation

The interpretation of the final ALSCAL mapping which follows draws on four sets of inputs:

- the outputs of the multivariate analysis of the correlation matrix
- secondary, non-bibliometric, sources of information
- a subject analysis of the citing literature
- personal knowledge of the field of information policy studies

It has been noted already that author cocitation studies do not assume, as is the case in some other areas of citation analysis, that the networks of citations between authors necessarily represent cognitive relationships. The assumption is rather that highly patterned cocitation data may be representative of intellectual and / or social structure in some unknown proportion. Citation practices are well understood to be motivated by concerns other than simply giving credit for related work. Weinstock (1971:19) lists some of the other influences shaping citation practice:

- paying homage to pioneers
- providing background reading
- criticising or correcting one's own work or the work of others
- substantiating claims or disputing priority claims of others
- altering researchers to forthcoming work
- providing leads to poorly disseminated, poorly indexed or uncited materials

Much has been written on the nature of academic disciplines in the educational literature, where the matter is of course of practical relevance in scoping and designing curricula. Some of the views expressed in the educational literature are highly polarised—between those who focus on epistemological considerations, presenting disciplines as "each characterised by its own body of concepts, methods and fundamental aims" (Toulmin, 1972:16) and those who see disciplines unequivocally as organised social groupings (Whitley, 1976 & 1984).

King & Brownell (1966) offer a wide-ranging and more balanced account of the nature of disciplines. They embrace several different aspects: a community of scholars, a network of communications, a tradition, a particular set of values and beliefs, a domain, a field of enquiry, and a conceptual structure. Perhaps the sharply polarised views of Toulmin and Whitley are unhelpful—Shinn, after studying academics working in mineral chemistry, solid-state physics, and

computerised vector analysis concluded that "the internal structure of the cognitive and social arrangements match" (Shinn, 1982: 222), a point echoed by Becher, another educationalist:

> *"It would seem, then, that the attitudes, activities and cognitive styles of groups of academics representing a particular discipline are closely bound up with the characteristics and structures of the knowledge domains with which such groups are professionally concerned. One could venture further to suggest that in the concept of a discipline the two are so inextricably linked and connected that it is unproductive to try to forge any sharp division between them. Even so, if one is to examine the nature of their interconnections, a distinction must be made - at least in theoretical terms - between forms of knowledge and knowledge communities"* (Becher, 1989:20).

These insights from the educational literature seem very pertinent to arriving at a robust interpretation of author cocitation data. Most authors of published ACA studies have acknowledged the ambiguous nature of their data, although it must be noted that the cognitive dimension has tended to receive almost exclusive emphasis in forming their interpretations. Clearly, any attempt to validate an interpretation based in whole or in part on forms of social organisation would require a separate qualitative investigation, perhaps along the lines of the semi-structured interview methodology developed by Becher (1989).

It is important in developing an interpretation to consider whether factors such as nationality, institutional affiliation, or employment sector might offer some clues as to the meaning of the patterns revealed by multivariate techniques. With these thoughts in mind, biographical information was collected on each author. The objective was to build up a thumbnail sketch of each individual, in terms of their primary research interests, professional and institutional affiliations, career patterns, and so on. A wide variety of information sources were scanned, including:

- entries in *Who's Who* type publications
- biographical information on book jackets, conference programmes, etc.
- personal World Wide Web pages
- possible authorship of entries in LIS or related encyclopaedias
- potted biographies associated with journal articles
- lists of academic prizewinners

Some basic biographical information is summarised in Table 8.8 overleaf. Of the 21 authors in Table 8.8, nine are (or were) academics at schools of librarianship or information science; one a legal academic; six policy specialists in government, independent think tanks or the private sector; three academic or national librarians; one private consultant; and the (former) director of a trade body. One striking conclusion which may be drawn immediately from this Table is the close correspondence between the groupings indicated by factor and cluster analysis and author nationality. All the authors associated with Factors 1 and 3 are North American; and with the exception of Victor Rosenberg and Toni Carbo Bearman (British-born and formerly employed at *INSPEC* before moving to the States) all the authors associated with Factors 2 and 4 are British. It is therefore possible to speculate that nationality is a key factor in determining group membership, although it is unlikely that this represents anything other than the surface of a rather deeper set of issues. Certainly, the mechanisms for research funding and the reward systems for researchers are very different, but then so is the fundamental nature of the information policy environment. It may well be that the information policy issues and concerns at national or regional level shape and dictate information policy research agendas[57].

[57] Analysis of the document test collection showed that 67.3 per cent of the articles were primarily concerned with policy developments at national or regional level; only 24.9 per cent with policy at the supranational level (the remaining 7.8 per cent were unassigned). Very few studies of an international comparative nature were recorded.

Table 8.8: Author biographical information

Author	Nationality	Period active(a)	Most recent corporate affiliation
ANTHONY	British	1980-87	Director, Aslib (retired)
BEARMAN	American	1975-95	Dean, Graduate School of Librarianship and Information Science, University of Pittsburgh
BORTNICK	American	1979-91	Analyst, Science Policy Research Division, Congressional Research Service, Library of Congress
BURGER	American	1978-93	Librarian, Area Studies Division, University of Illinois at Urbana-Champaign
CASE	American	1981-96	Professor, Department of Library and Information Science, University of Kentucky
CHARTRAND	American	1976-87	Senior Fellow, Congressional Research Service, Library of Congress
CRONIN	British	1979-96	Dean, Department of Information and Library Studies, University of Indianapolis
DURRANCE	American	1979-96	Associate Dean, School of Librarianship and Information Science, University of Michigan at Ann Arbor
FLAHERTY	Canadian	1974-92	Professor of Law, University of Western Ontario
HERNON	American	1973-96	Professor, Graduate School of Librarianship and Information Science, Simmons College, Boston
HILL	British	1976-95	Director, British Library Science Reference Information Service (retired)
KATZ	American	1980-94	Bell Communications Research Ltd, New Jersey
MARTYN	British	1972-92	Senior Fellow, Centre for Communication and Information Studies, University of Westminster (retired)
MCCLURE	American	1974-97	Professor, School of Information Studies, Syracuse University, New York
MOORE	British	1977-97	Senior Fellow, Policy Studies Institute
MOREHEAD	American	1973-95	Professor, Albany School of Library and Information Science, New York
MORTON	American	1973-96	Deputy Head (Reference), Montana State University Library
OPPENHEIM	British	1975-96	Director, International Electronic Library Research, de Montfort University, UK
RELYEA	American	1973-95	Specialist in American National Government, Congressional Research Service, Library of Congress
ROSENBERG	American	1980-96	Associate Professor of Information and Library Studies, University of Michigan at Ann Arbor
SPREHE	American	1974-97	Director, Sprehe Information Management Associates

(a) Year of first and most recent publication in *Social SciSearch* (as first author).

Given the difficulties in developing a social interpretation of this data, the remaining discussion will centre on the identification of possible subject specialties within information policy. As a tool to guide this interpretation, recourse was made to the online sets created during the data collection phase. In order to determine cocitation counts for all possible pairs of authors, 210 sets were created in *Social SciSearch.* These sets were combined (using OR logic) and the titles, abstracts and any indexing terms printed. Then, *for each individual author,* the works citing him or her within this universe were classified according to the faceted schema developed in Chapter 3. The objective of this step was to determine the underlying subject perceptions of the citing authors.

Figure 8.10: Subject analysis of the citing literature

General articles on information policy	
Burger	*11 - Theoretical aspects of information policy*
Chartrand	*11 - Theoretical aspects of information policy*
Hill	*12 - National and international information policies*
Information infrastructure policies	
Moore	*20 - Information infrastructure policies*
Anthony	*21 - Research & development (including STM information policies)*
Rosenberg	*21 - Research & development (including STM information policies)*
Bearman	*22 - Libraries, archives and public records*
Information management in government	
Hernon	*30 - Information management in government*
McClure	*30 - Information management in government*
Sprehe	*32 - IRM in government: policies and practice*
Case	*33 - Government information, clearinghouses and dissemination*
Durrance	*33 - Government information, clearinghouses and dissemination*
Morehead	*33 - Government information, clearinghouses and dissemination*
Morton	*33 - Government information, clearinghouses and dissemination*
Information access and control	
Flaherty	*42 - Confidentiality and personal privacy*
Katz	*42 - Confidentiality and personal privacy*
Bortnick	*43 - Information control on grounds of national security*
Relyea	*43 - Information control on grounds of national security*
Information industry policies	
Cronin	*50 - Information industry policies*
Martyn	*50 - Information industry policies*
Oppenheim	*53 - Regulation of information industry and markets*

The results are summarised in Figure 8.10, which indicates the modal broad subject category into which the citing documents fell. This is admittedly a crude procedure, but the intention was simply to develop a *Gestalt* impression of the

nature of the citing documents, not a detailed subject analysis. The results tend to support a subject-based interpretation of the groupings revealed by the factor analysis, clustering and multidimensional scaling procedures.

Figure 8.11: Subject specialties within the information policy domain

Specialty A: Information management in government
Hernon, McClure, Morehead, Morton, Sprehe
Federal information resource management; government publications; depository libraries; legal deposit; charging and pricing of official information; printing and dissemination.

Specialty B: Information protectionism
Bortnick, Relyea
Information policy and national security controls; official secrecy; Congressional oversight; transborder data flows; international information issues.

Specialty C: Public access to information
Burger, Case, Chartrand, Durrance
Access to government information; public information needs; evaluation of Federal and State library and information services; social construction of information technologies.

Specialty D: National & international frameworks
Hill
National and international information policies and strategies; national libraries; international collaborations and agreements; scientific and technical information.

Specialty E: Scientific & technical information policy
Anthony, Rosenberg
Scientific, technical and medical information policy; research & development; technology transfer; online information resources in science, technology and medicine.

Specialty F: Information infrastructure & regulation
Bearman, Cronin, Martyn, Moore, Oppenheim
Information industry and markets; political economy of information; manpower and training; public-private synergy; copyright and intellectual property; information law & regulation.

Specialty G: Social implications of ICTs
Flaherty, Katz
Information technology and privacy; data surveillance; freedom of information and expression; media law and controls.

Most of the authors associated with Government Information in the factor analysis appear under the main heading of Information Management in Government. Similarly, the other three factors resolve fairly neatly into the classification scheme. Taken together with the information about cluster group

membership (Figure 8.8), this analysis suggested that some finer subdivision of the Government Information factor might be possible.

After consideration of all the available sources of evidence, including the two-dimensional ALSCAL plot, the authors were partitioned into seven subject groupings or 'schools'. These are shown in Figure 8.11 together with a brief scope note. The first three groupings: Information Management in Government; Information Protectionism; and Public Access to Information can be regarded as sub-sets of a macrocluster defined by the Government Information factor in the principal components analysis. These groupings are projected onto the ALSCAL plot in Figure 8.16 (p.185). This shows Michael Hill (National and International Policy Frameworks) as the most central figure, and arranges the North American government information specialists to the left and the predominantly British information industry and infrastructure policy specialists to the right, in close proximity to the more central scientific and technical information policy authors. On this basis of this interpretation, the horizontal axis appears to reveal a polarisation between the State (left) and the broader information industries (right). The vertical axis is more difficult to interpret given how sparsely populated the authors are below the origin in the final map. It is notable that, unlike the other authors, David Flaherty and James Katz come from outside the LIS tradition (Flaherty is a law professor, Katz a media and communications policy specialist employed at Bell Communications). It may be that the *x* axis marks a boundary with non-LIS research traditions, or possibly that it distinguishes between conceptual or methodological approaches[58].

[58] In the early stages of the research, a map of all 34 candidate authors was created and subsequently rejected since 13 of the authors failed to meet the threshold criteria (see Figure 8.5). A notable feature of this early map was that it tended to place non-LIS authors (as defined by institutional affiliation and propensity to publish in non-LIS journals): Braman, Flaherty, Katz, Irwin and Sauvant, below the horizontal axis and LIS authors above. This suggests that the horizontal axis expresses (unidentified) factors which relate to existing disciplinary perspectives.

8.7 Validation

The work which has been presented so far suggests that authors in the field of information policy may be resolved into a smaller number of clans on the basis of the highly patterned nature of their cocitations. The maps which result from the analysis of communication artifacts are interesting, but what do they tell us about the nature of the scholarly community of information policy? White & McCain remind us that:

> *"While each person in a cocitation analysis is technically a body of writings rather than a person, it is evident, when viewing the maps and clusters, that living authors grouped as intellectually related oeuvres are often also socially related as persons. Part of the fascination of author maps lies in seeing such relationships emerge from the automated processing of citations: repeatedly, authors who are proximate on the map have not only subject matter and method in common but also collaborative ties. (Other possibilities are ties of language, period, nation, or ideology). The maps are thus reminders of the interplay of social and intellectual structure and speak to questions about it raised by sociologists, such as Crane and Mullins. This part of cocited author analysis needs more systematic development"* (White & McCain, 1989:148).

These thoughts prompted the idea that it might be possible to validate and even to extend the interpretation of the author cocitation data by means of a simple postal questionnaire to the authors involved. The purpose of the questionnaire was two-fold: firstly to collect information about any existing social and collaborative ties between authors and, secondly, to see whether the authors agreed with how they had been allocated to the subject clusters shown in Figure 8.11. The final questionnaire and accompanying documentation are attached as Appendix F.

The survey generated considerable interest and a very high response:

Table 8.9: Postal questionnaire: response rate

Questionnaires sent	21
Useable responses	19
Returned, address incorrect	2
Valid response rate (%)	100.0

The questionnaire asked four simple questions. Respondents were asked:

- whether they recognised the names of six other authors in the study and whether they could (a) put a face to that name, and (b) whether they were reasonably familiar with the writings of that author

- to tick a series of boxes which represent the existence of some form of social or collaborative tie (such as maintaining regular correspondence or working together on a research proposal) with each of the six authors

- to identify which of the six authors they perceived to be their `closest' intellectual relations and which the `most distant'

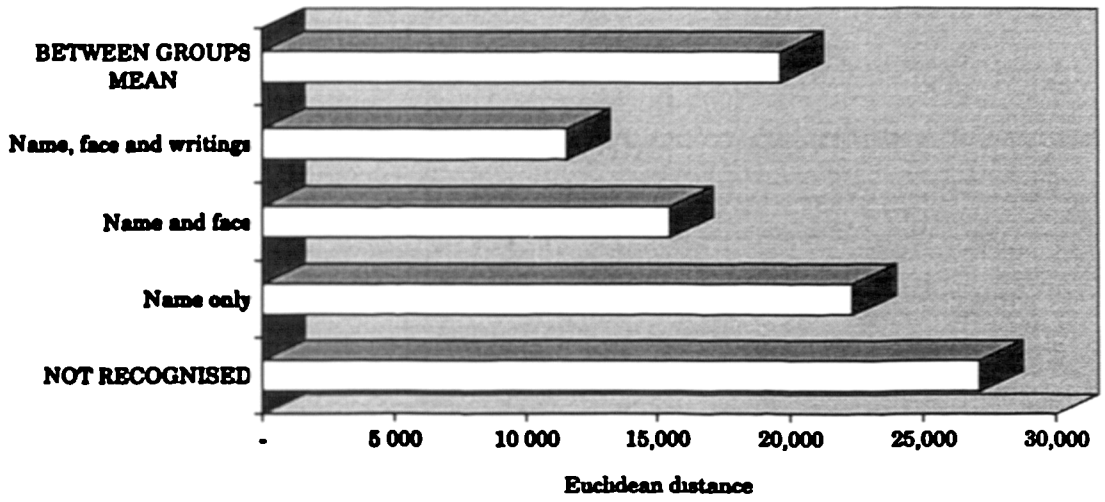
- to indicate whether they agreed with their allocation to a subject cluster

The questionnaire was personalised in each case in such a way that the six authors included were different in each case. The six authors were selected on the basis of their correlation profiles with the data subject—the three highest Pearson coefficients and the three lowest (subject to the qualification that at least one cocitation must be present in each case). The respondents were presented with the authors in alphabetical order and were not shown the ALSCAL map, nor any other indication of cocitation structure. Although only 19 questionnaires were available for analysis, each provided data for six author pairs, a total of 114 pairs in all.

8.7.1 Recognition

The results of Question 1 are both striking and statistically significant[59]. Figure 8.12 shows the mean Euclidean distance (taken from the *SPSS* proximity's matrix) for the whole sample and broken down by degree of recognition.

Figure 8.12: Recognition and mean Euclidean distance



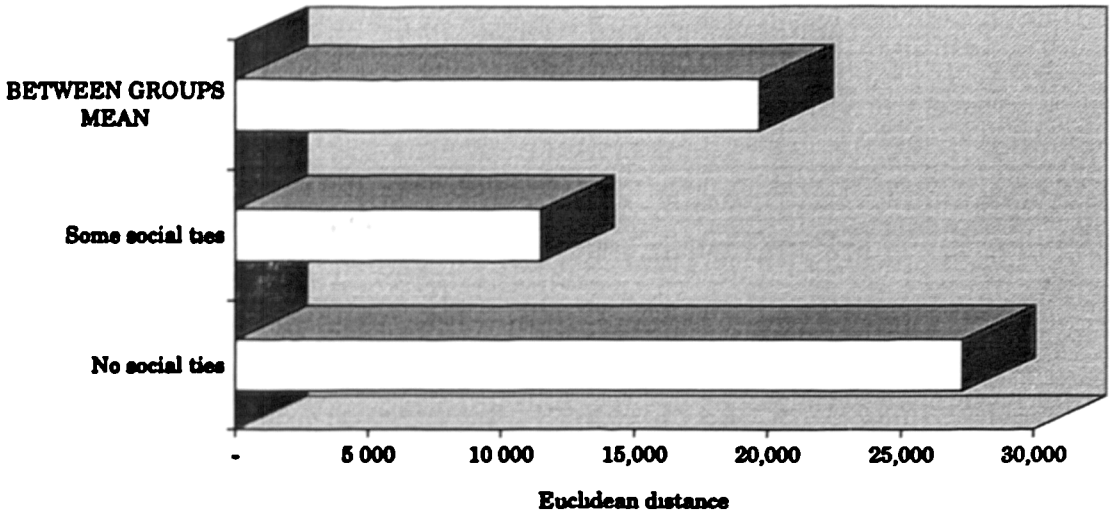
These results appear to support the observation of White & Griffith that "one writer, upon seeing the map, said he knew everyone about two-thirds across the map and then no one" (White & Griffith, 1981: 171).

8.7.2 Social and collaborative ties

The results of Question 2 were also highly significant[60]. Figure 8.13 overleaf shows the mean Euclidean distance associated with two subsets: those author pairs where no social or collaborative ties exist, and those indicating ties of any kind.

[59] $F = 38.9$, d.f. = 3, $p < 0.001$
[60] $F = 158.8$, d.f. = 1, $p < 0.001$

Figure 8.13: Social and collaborative ties and mean Euclidean distance



One direct implication of these findings is that authors grouped together by the clustering routines are much more likely to share social and collaborative ties than those placed further apart. This strongly suggests that the ALSCAL map is indicative of some form of social as well as cognitive structure. This point is amplified in Figure 8.14, a three-way contingency table:

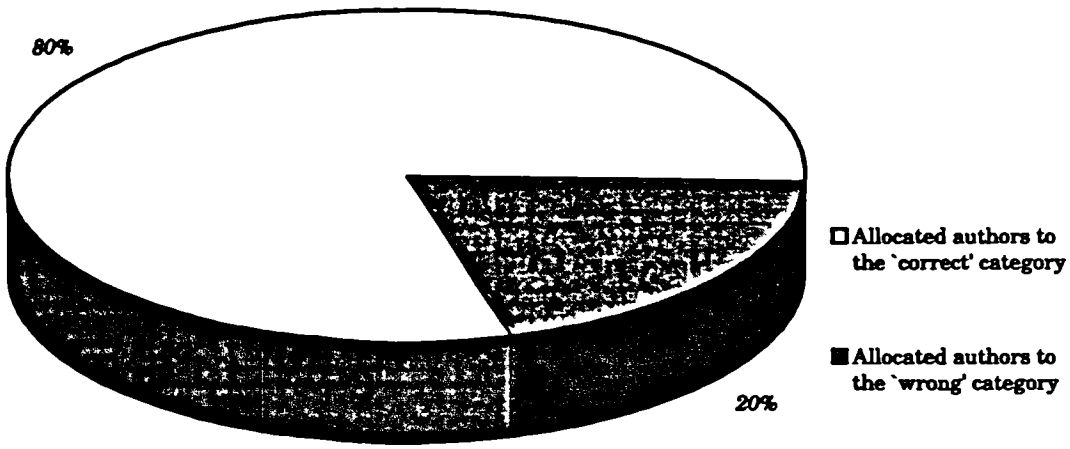
Figure 8.14: Recognition, social ties and proximity

Three closest neighbours			**Three furthest neighbours**		
Recognise name?			*Recognise name?*		
YES		NO	YES		NO
Any social ties?			*Any social ties?*		
YES	NO		YES	NO	
49	1	7	5	6	46

8.7.3 Intellectual distance

In Question 3, respondents were asked whether they could say which three authors were their closest or most distant `intellectual relations'. This question proved to be very difficult for many of the respondents—of the 114 author pairs sampled, 55 `don't knows' were returned, chiefly in relation to the `most distant' category where, often, the names of the authors were unfamiliar. The results still tend to support the credibility of the ALSCAL mapping.

Figure 8.15: Intellectual distance



8.7.4 Subject clusters

In Question 4, the respondents were given the opportunity to assign themselves to one of the subject specialties inferred by the present author from the ALSCAL mapping:

Table 8.10: Subject clusters

	n
Agreed with cluster allocation as first choice	11
Agreed with cluster allocation as second choice	4
Chose an adjoining cluster instead	2
Chose a non-adjoining cluster	2
Total	*19*

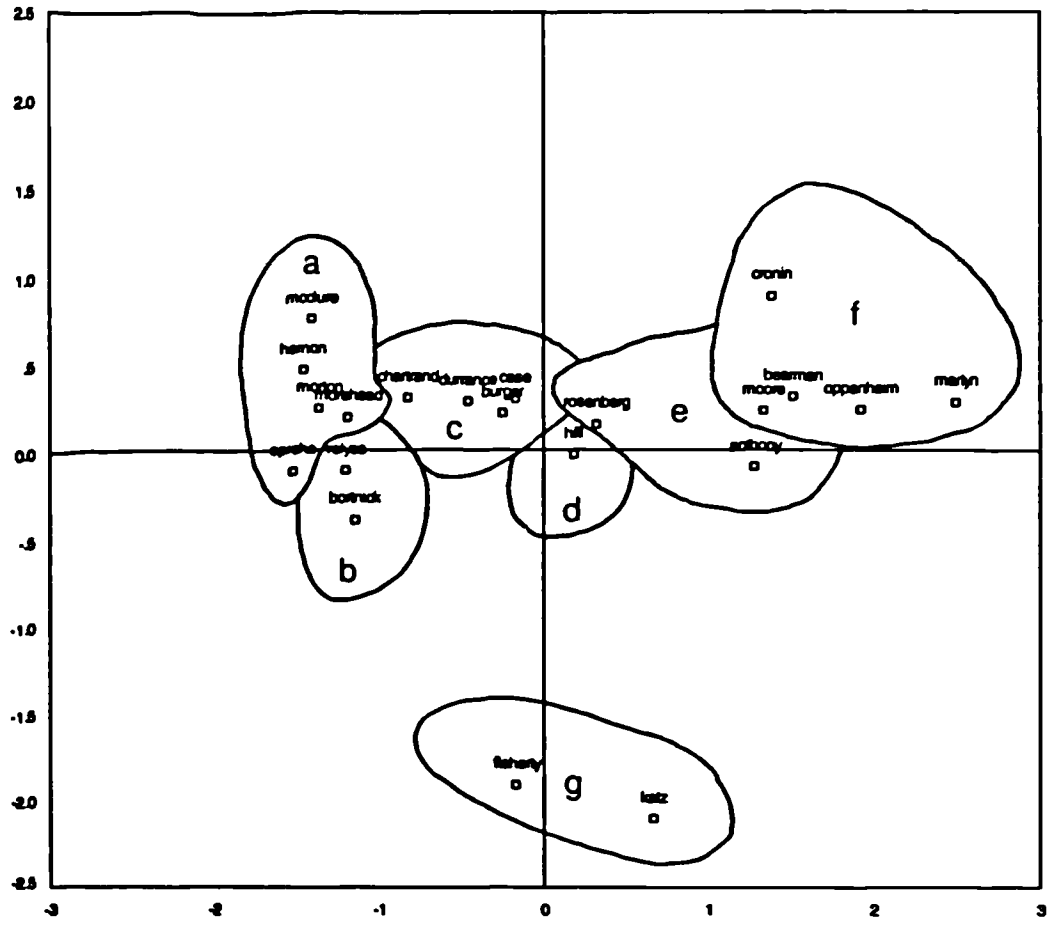
On reflection, this question was problematic for a number of reasons. While roughly half of the respondents agreed with their cluster allocation as their first choice, one respondent noted that that the specialty areas have considerable overlap (especially D and E). This view was clearly shared by others who were frustrated by the constrained choice available to capture what, in many cases, was a broadly-based portfolio of interests. In several cases, authors allocated themselves to the broadest and most fundamental category available (D - National and International Information Policy Frameworks) but then chose either the 'right' category or an adjoining cluster as their second choice.

The respondents' own views on where they feel they should be located need to be treated with a great deal of caution, not least because there is no evidence that the papers which cocited them are in any way representative of their total published output or current research interests.

8.8 Conclusions

It is contested here that the empirically-derived groupings of authors presented in Figure 8.16 on the next page offer unique insights into the intellectual and social structure of the information policy domain. The groupings resulting from the various multivariate analyses were shown to be highly consistent both with one another and with an independent analysis of the subject content of the citing articles. The results have face value and considerable intuitive appeal. This is consistent with the basic premise of author cocitation studies that the results draw upon the collective citation practices of a potentially very large population of citing authors.

Figure 8.16 Final ACA interpretation



Despite these remarks, this study is limited in a number of respects, some of which are integral to the author cocitation methodology itself. The sins of omission and commission which may impact on the selection of authors to study have already been mentioned. McCain (1990) argues that if a final map is to be truly representative it is important to identify authors whose work is not found in the standard texts and reviews—these authors are "likely to represent newly prominent or non-traditional approaches to the subject matter ... social science areas with strong political factions or schools of thought may be particularly problematic" (McCain, 1990:434). There is certainly a danger that in their search for `objective' criteria for selecting authors, cocitation researchers will tend towards a rather conservative view of the field of studies being considered. This is likely to be reinforced by practical considerations—the need to identify authors who are highly cited and who have already established their reputations. Another practical consideration in author cocitation work is the

almost axiomatic choice of *ISI* databases as the data source: this means that scope is circumscribed entirely by *ISI* editorial policy. As a result, "ACA can be nothing but biased in favour of Anglo-American research" (Kärki, 1996).

One serious limitation of the present study is the relatively small number of authors appearing in the final mapping—many published studies have typically included 35 or more authors. This may account for the very sparsely populated lower quadrants of the ALSCAL plot (alternatively, the authors selected may not be truly representative of the field). There was a very clear trade-off in this case between increasing the number of authors (thus generating maps at higher levels of resolution) while maintaining a high degree of data integrity and stability. The threshold criteria (based on the work of White & Griffith, 1982) which were used to screen information policy authors for inclusion in the study proved to be very stringent in the context of a literature with comparatively low rates of cocitation. McCain (1990) suggests that studies based on mean cocitation rates as low as four may still yield *prima facie* acceptable results, however.

Another related problem stems from the extended time-frame of the present study, covering the period 1972-January 1997 (although roughly 60 per cent of the citing documents retrieved were published after 1988). Most cocitation studies have a more restricted time-frame, typically four or five years—this would however have been impracticable given the low rates of cocitation in the information policy literature. Ideally, if higher rates of cocitation had been encountered, the present study would have attempted to generate a longitudinal series of maps showing how the relations between authors had changed over time. The fact remains that the stability of the data in the present study over time is unknown and therefore a cause for concern. The longer the career of a given author, the more likely it is that his or her interests may have shifted.

The ambiguous meaning of cocitation data and the possible distortions introduced by idiosyncratic or erratic citation practice have already been touched upon. At the level of an individual citing author, this is a valid criticism; however, author cocitation analysis operates at the level of a *population of citing*

authors (181 in the case of the present study) and so the credibility of the technique[61] must be judged at the level of citation practices across the field as a whole.

In conclusion, 'Is information policy is a single coherent field of study'. The evidence presented in this Chapter would suggest 'probably not'. The high degree of independence of the factors identified in the principal components analysis suggests that information policy scholarship is not (yet?) contributing to a coherent body of knowledge, although some weak co-dependence was demonstrated between two of the factors involved (Scientific and Technical Information Policy and Information Infrastructure Policy). The apparently low rates of cocitation observed in this study are low only in relation to published author cocitation studies, raising the question of whether the networks of cocitation in the information policy area really are attenuated or whether previous workers have selectively focused on 'richer' sources of such data. On the other hand, the high levels of observed connectedness between the authors, mostly over 70 per cent and the incidence of highly positive and significant correlation profiles suggests that the cocitation data is highly structured and the resulting maps not merely an artifact of the (admittedly elaborate) data preparation and analytical techniques used.

The evidence here strongly supports the notion that there are two major schools of information policy scholarship, the contours of which are determined by political and legislative factors: one, a US school, responding to the policy issues flowing from a very highly developed government information infrastructure[62], the other a predominantly British school with a broader set of concerns in the political economy of information.

---

[61] See, for example, Edge (1977) who argues that citation is a 'relatively trivial behaviour' and that cocitation studies tend to accumulate and average and thus 'destroy' the rich evidence provided by individual variations.

[62] Consider, for example, the American culture of openness and its highly developed public information policy environment (the First Amendment; 'Government in the Sunshine' and Freedom of Information legislation; Paperwork Reduction Act, and so on), and the activities of the Office of Management & Budget (OMB) and the Office for Technology Assessment (OTA) in the information sector. It would be difficult to draw parallels with the UK in any of these areas.

Chapter 9: Content analysis

> *"Read not to contradict and confute, nor to believe and take for granted, nor to find talk and discourse, but to weigh and consider"*[63]

9.1 Introduction

So far, the analysis of the document test collection has been based on elements which were either already present in the records downloaded from the *Social Science Citation Index*, or on elements which could be derived with little intellectual effort. The content analysis reported here extends that earlier work by investigating the distribution of various indicators of content. These indicators are qualitative and judgmental. They codify such factors as the information policy topics under investigation; the aims, scope and intentions of published articles; the research strategies and methods employed; and the nature of corporate authorship. The content analysis variables used here are listed below as Figure 9.1:

Figure 9.1: Content analysis variables

ACAD	*Academic or practitioner*
AIMS	*Main objective of article*
FOCS	*Analytical focus*
INS2	*Narrow author status*
INTN	*Intent of article*
MTHD	*Research method*
ROWL	*Information policy research strategies*
SBJC	*Subject analysis of article*
SCAL	*Information policy scale*
SCOP	*Scope of article*
STGE	*Stage in policy life-cycle*
TYPE	*Research-based or opinion article*

Content analysis shares many of the characteristics positively associated with bibliometrics, in that it is (a) unobtrusive; (b) easily replicable; and (c) permits

[63] Francis Bacon, *Essayes: 50—Of Studies* (1597).

the analysis of large volumes of qualitative data. Somewhat surprisingly, content analysis does not appear to have been widely used as a tool in bibliometric studies[64], despite its potential for illuminating important aspects of communication (e.g. meaning) which tend to be neglected within the more abstract framework of bibliometrics[65].

Content analysis is a broad term that has come mean different things to different researchers[66]. Sheppard & Bawden (1997) point out that content analysis embraces a variety of techniques and approaches, emphasising qualitative as well as quantitative analysis, and the inclusion of latent as well as manifest content. Essentially, content analysis is a type of textual analysis, the aim being to identify meaning buried in the text. Content analysis has been defined as:

> *"a research technique for the objective, systematic, and quantitative description of the manifest content of communication"* (Berelson, 1952:18).

The approach adopted in this study is purely quantitative: units of analysis (individual articles) are enumerated in terms of their frequency of occurrence, disaggregated according to their membership of the categories shown in Figure 9.1. The decision to use individual full text articles as the unit of analysis meant that rough justice had to be applied in some cases since papers could only be assigned to a single exclusive category. Given the nature of the categories involved, many of which refer to rather abstract notions of underlying strategies and motivations, subjective judgements were required to infer meaning that was in many cases latent rather than explicit. Only papers which were judged to be research-based were included in this part of the study.

---

[64] See, however, Cronin and others (1997) who incorporate content analysis in an excellent study of the womens' studies literature.

[65] Pierce offers some interesting thoughts on content analysis and bibliometrics and argues that the two techniques may be much more closely related than has been generally recognised (Pierce, 1990).

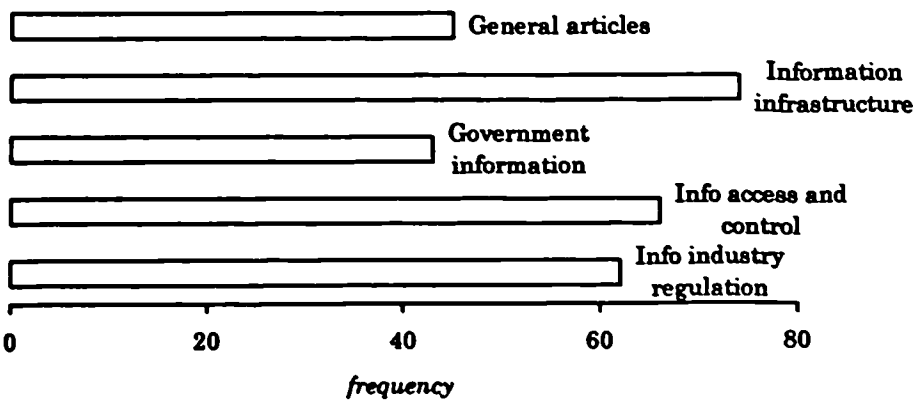
[66] The reader is directed to Krippendorff (1980) for a comprehensive text on content analysis, dealing with both its theoretical and practical aspects.

Because of the subjective nature of the way that articles were assigned to categories, a number of safeguards were taken to try to ensure a reasonable level of stability in recording the data. In all cases, the original full-text article was consulted since it was often impossible to infer sufficient meaning from the author abstracts supplied by *ISI*. Data recording took place over a concentrated two-week period in February 1997. Three months later, fifty articles were selected at random and recoded for each of the 12 categories in the content analysis frame—but without consulting the coding decisions that had been made earlier. When compared with the first attempt at coding, a high degree of stability was obtained—the two coding exercises agreed on 574/600 (95.7 per cent) of occasions. The most problematic assignments were to the categories which represented the scope (SCOP) and intent (INTN) of the articles, which were also those variables where it was least usual to find a manifest expression in the text.

9.2 Information policy topics

This section considers how information policy topics are distributed across the test collection as a whole and then explores the influence of author region[67], journal category, and time span on this distribution. Figure 9.2 shows that the bibliography items as a whole are fairly evenly spread across the broad headings of the subject nomenclature:

Figure 9.2: Broad subject headings (n = 771)



[67] 'Author region' reflects the geographic location of the first named author, as indicated by his/her corporate address.

Bivariate analyses show, however, that there are some striking differences in the ways that broad subject headings are distributed. Table 9.1, for instance, shows a statistically significant relationship between broad subject and author region:

Table 9.1: Cross-tabulation: broad subject by author region

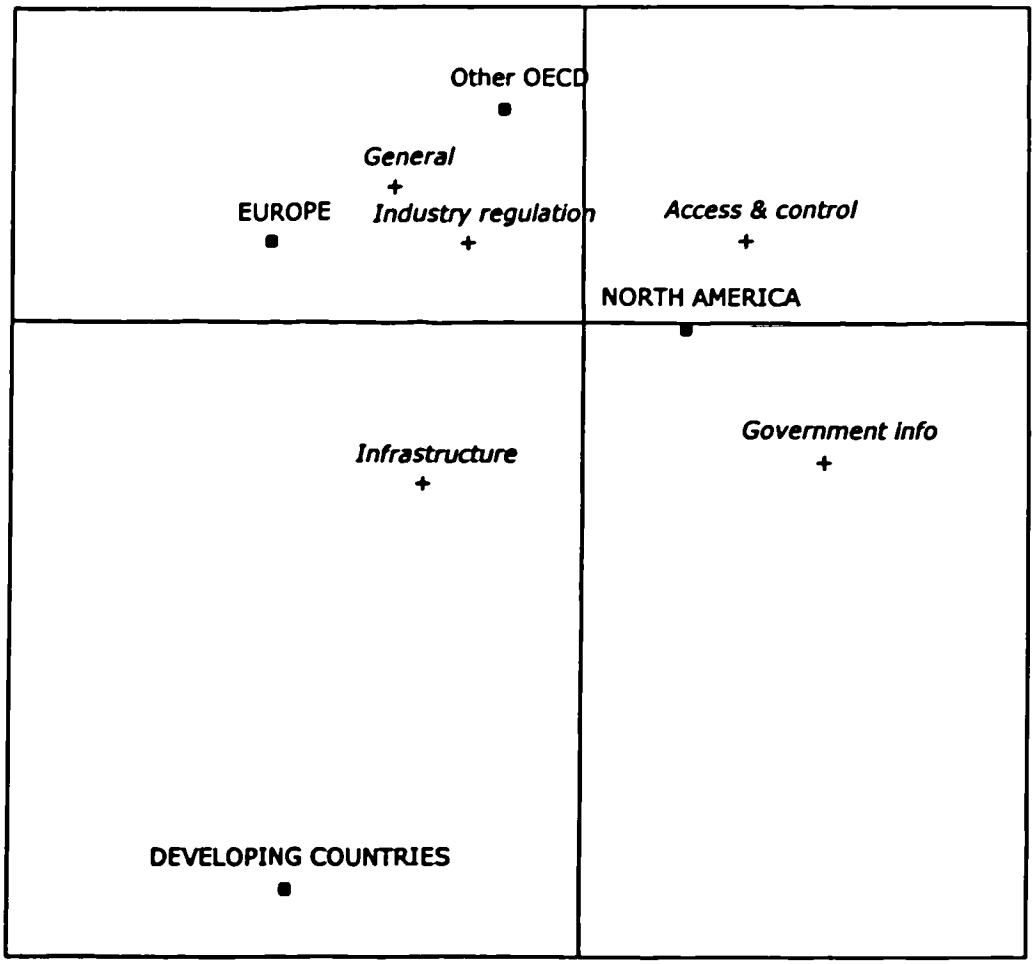
	General articles	Infra-structure	Government information	Access & control	Industry regulation
North America	65	111	104	197	85
	-3.7	*-4.0*	*5.0*	*4.7*	*-2.4*
European Union	37	47	4	24	37
	4.0	*2.6*	*-4.6*	*-4.2*	*2.8*
Other OECD	6	7	1	12	6
	0.7	*-0.2*	*-1.9*	*0.9*	*0.2*
Developing countries	3	16	4	1	4
	-0.6	*4.3*	*-0.1*	*-3.1*	*-0.4*

The data in Table 9.1 are frequencies (normal text) and standard residuals[68] (in bold italics). The null hypothesis, H_0, that broad subject and author region are independent, must be firmly rejected[69]. Initial examination of Table 9.1 shows that the topic headings 'Government Information' and 'Information Access and Control' are more heavily represented in the writings of North American authors than would be expected if broad subject and author region were independent. Similarly, it reveals a propensity among European authors to write general articles and papers on information infrastructure topics. A less formal mapping of the two variables is presented as Figure 9.3 overleaf using a graphical technique known as correspondence analysis (ANACOR).

[68] 'Residuals' are an expression of the difference between observed and expected frequencies. In order to make the table easier to interpret, residual values are presented here using an estimate of their standard error, expressed in standard deviation units above or below the mean. A positive value indicates that there are more cases in a cell than there would be if the row and column variables were independent.

[69] Chi-square = 87.1, d.f. = 12, $p < 0.001$; Cramer's V = 0.19, $p < 0.001$

Figure 9.3: ANACOR: broad subject by author region



Throughout this Chapter, use is made of correspondence analysis as a complementary technique to formal non-parametric tests such as Chi-square and Cramers' V. Correspondence analysis computes row and column scores from the contingency table and then generates plots based on the joint scores. It then represents the relationship between two nominal variables as a scatter diagram (as in Figure 9.3). It is a particularly useful tool for visualising the nature of bivariate relationships in cases where more than two categories are involved and where there is supporting evidence that the two variables are not independent.

Nagpaul & Sharma, in a valuable technical paper, argue that correspondence analysis allows the user to construct typologies which are "useful for further discussion, research and pedagogy" (Nagpaul & Sharma, 1995:265). They offer some useful keys for interpreting the resulting scatter diagrams:

1. The centre of gravity located at the origin of the axes corresponds to the average profiles of both sets of points (i.e. broad subject and author region in Figure 9.3). Elements closest to the centre show little differentiation with respect to the second variable; thus it can be seen that the subject profile of North American authors is much closer to the average than is the profile of writers in the developing countries.

2. The smaller the distance between two categories of the *same* variable, the closer their profile with respect to the second variable. The regional profiles of 'General Articles' and 'Information Industry Regulation' are thus seen to be more similar than, say, 'Information Industry Regulation' and 'Government Information'.

3. The smaller the distance between two categories from *different* variables, the more intense their interrelationship; conversely elements which are far apart may be assumed to be more or less independent: as in the case, say, of 'Information Industry Regulation' and 'Developing Countries'.

It is worth pointing out that such inferences should only be drawn where the ANACOR mapping accounts for a high proportion of the total variance[70].

The results above show that there is a significant relationship between broad subject and author region, with European authors tending to write more general articles and more papers on information infrastructure issues than might have been expected, while North American writings appear to be more concerned with issues concerning the exploitation and control of access to government information. This finding parallels the results of the author cocitation analysis (Chapter 8) which also concluded that the subject profiles of the North American and European authors included in the study were highly differentiated. The relatively low numbers of papers from authors in other OECD and developing countries suggest that further attempts at interpretation would be unwise.

[70] The two-dimensional mapping in Figure 9.3 accounts for 98.3% of the total variance on two factorial axes, x (79.0%) and y (19.3%).

Not surprisingly, the distribution of broad subject headings also exhibits a statistically significant relationship with *ISI* journal category. This finding accords with the results of the discriminant analysis carried out earlier (Table 7.4, p.134) which showed that journal category was a good predictor of Ward cluster membership.

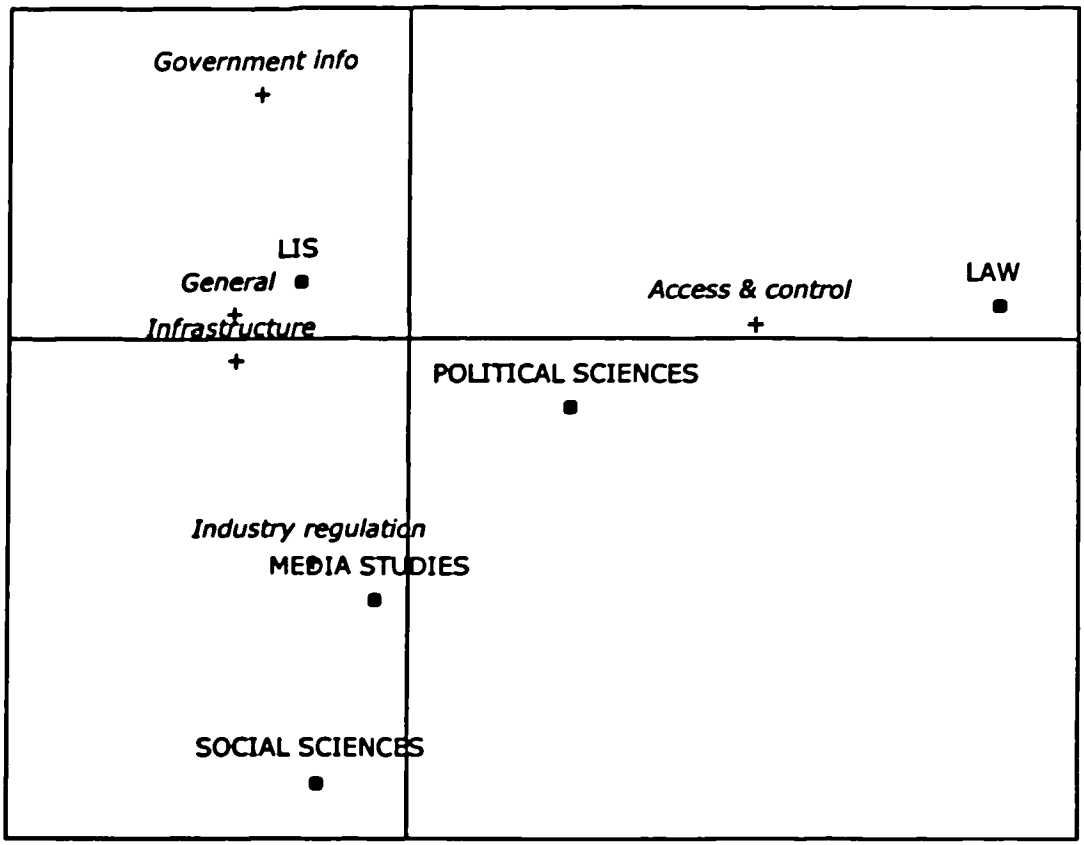
Table 9.2 and Figure 9.4 show that broad information policy subject headings are distributed asymmetrically across *ISI* journal categories:

Table 9.2: Cross-tabulation: broad subject by journal category[71]

	General articles	Infra-structure	Gov'mnt information	Access & control	Industry regulation
LIS	93	147	100	110	90
	3.4	*3.8*	*4.6*	*-9.2*	*-0.5*
Media studies	4	8	1	7	6
	0.1	*0.9*	*-1.6*	*-0.4*	*0.8*
Law	2	2	3	79	6
	-3.6	*-5.1*	*-3.3*	*12.3*	*-2.9*
Political sciences	4	10	8	28	13
	-1.9	*-1.5*	*-0.5*	*2.5*	*0.8*
Social sciences	8	14	1	10	17
	0.3	*0.8*	*-2.6*	*-1.6*	*3.3*

[71] Chi-square = 193.0, d.f. = 16, $p < 0.001$; Cramer's V = 0.25, $p < 0.001$

Figure 9.4: ANACOR: broad subject by journal category[72]



On the evidence of these findings, some interesting preliminary conclusions emerge about the collective decisions of information policy authors regarding where they publish. LIS journal titles tend to carry more articles of a general nature and more papers on government information and information infrastructure issues than would be expected if journal category and broad subject heading were independent variables. Papers on information access and control tend to be written for law or political science journals, and articles on information industry regulation for media studies titles. In terms of its distance from the origin of the two axes, the LIS journal category exhibits less differentiation in broad subject terms than any other journal category, except possibly political science. Law titles and social science journals are the most strongly differentiated categories by subject. This finding needs to be moderated, however, by the relatively high frequencies associated with the LIS journal category; a factor which will tend to push LIS towards the centre of gravity of the map.

[72] This two-dimensional mapping accounts for 98.6% of the total variance on two factorial axes, x (88.3%) and y (10.3%).

Table 9.3: Cross-tabulation: narrow subject by journal category[73]

	LIS	Media studies	Law	Political sciences	Social sciences
General articles					
Theoretical papers	33	4	1	1	7
	0.3	*2.1*	*-2.1*	*-1.5*	*2.5*
National information policy	60	0	1	3	1
	4.1	*-1.6*	*-2.7*	*-1.1*	*-1.7*
Information infrastructure					
Scientific information policy	29	0	0	4	2
	1.7	*-1.1*	*-2.2*	*0.7*	*-0.2*
National and public libraries	49	0	0	1	0
	4.5	*-1.4*	*-2.7*	*-1.6*	*-1.9*
Telecommunications policy	52	8	2	2	4
	1.2	*4.0*	*-2.4*	*-1.6*	*-0.2*
Info technology policies	17	0	0	3	8
	-1.1	*-1.0*	*-2.0*	*0.5*	*4.8*
Government information					
Information acquisition	12	0	0	0	0
	2.3	*-0.7*	*-1.3*	*-1.0*	*-0.9*
IRM policy and practice	40	0	1	7	0
	2.1	*-1.3*	*-2.2*	*1.7*	*-1.9*
Dissemination policies	48	1	2	1	1
	3.4	*-0.6*	*-1.9*	*-1.7*	*-1.4*
Information access and control					
Freedom of information	49	3	54	16	5
	-8.5	*-0.7*	*11.6*	*2.0*	*-1.3*
Confidentiality and privacy	23	3	17	6	4
	-4.4	*1.0*	*4.7*	*0.9*	*0.3*
National security	38	1	8	6	1
	0.1	*-0.6*	*0.7*	*0.8*	*-1.4*
Information industry regulation					
Standards and protocols	4	0	0	1	1
	-0.2	*-0.5*	*-0.9*	*0.8*	*1.0*
Copyright and IPRs	14	4	4	2	6
	-2.9	*3.1*	*0.2*	*-0.3*	*3.1*
Industry / market regulation	40	1	0	1	1
	3.4	*-0.4*	*-2.5*	*-1.4*	*-1.1*
Trade in information services	15	1	2	6	6
	-2.4	*0.0*	*-0.9*	*2.4*	*3.1*
Public-private synergy	17	0	0	3	3
	0.4	*-0.9*	*-1.8*	*0.9*	*1.3*

The distribution of narrow subject headings by journal category shown above as Table 9.3, while not statistically significant, tends to support the conclusion that there are differences in the kinds of information policy papers accepted for publication by different types of journal. The differences in profile between LIS journals and titles in law or the social sciences are once again rather striking.

[73] Chi-square = 337.8, d.f. = 64, $p < 0.001$ (NB 65% of cells have an expected frequency of less than 5, so it is not advisable to reject H_0); Cramer's V = 0.33, $p < 0.001$

Finally, a significant relationship between broad subject and age of publication is evident (Table 9.4). The division of the test collection into two time spans: 1972-1985 and 1986-1997 is arbitrary although it does permit some broad conclusions to be made in relation to a seminal paper by Trauth (1986), one of the very few papers to consider the nature of the information policy literature.

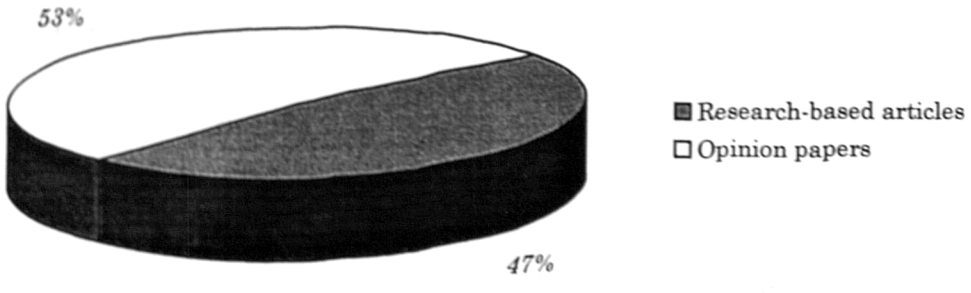
Table 9.4: Cross-tabulation: broad subject by time span[74]

	General articles	Infra-structure	Government information	Access & control	Industry regulation
Before 1986	26	35	15	122	21
	-1.3	*-3.1*	*-3.9*	*9.6*	*-3.5*
1986 forward	85	146	98	112	111
	1.3	*3.1*	*3.9*	*-9.6*	*3.5*

Table 9.4 shows that the distribution of broad topics covered in the information policy serials literature has undergone a major shift since 1985; the biggest change being a significant drop in the proportion of published papers dealing with information access and control issues. A similar conclusion was drawn earlier (Figure 7.3, p.140) when a clustering of the document test collection isolated a sharply-defined cluster of legal practitioners and academics writing almost exclusively on freedom of information, confidentiality and privacy issues arising out of topical policy events in the USA. More recent (1986-) papers represented in the document test collection are more likely to cover information infrastructure, information industry regulation or government information issues than earlier (pre-1986) papers.

9.3 Research methods

Figure 9.5: Research and opinion articles (n = 771)

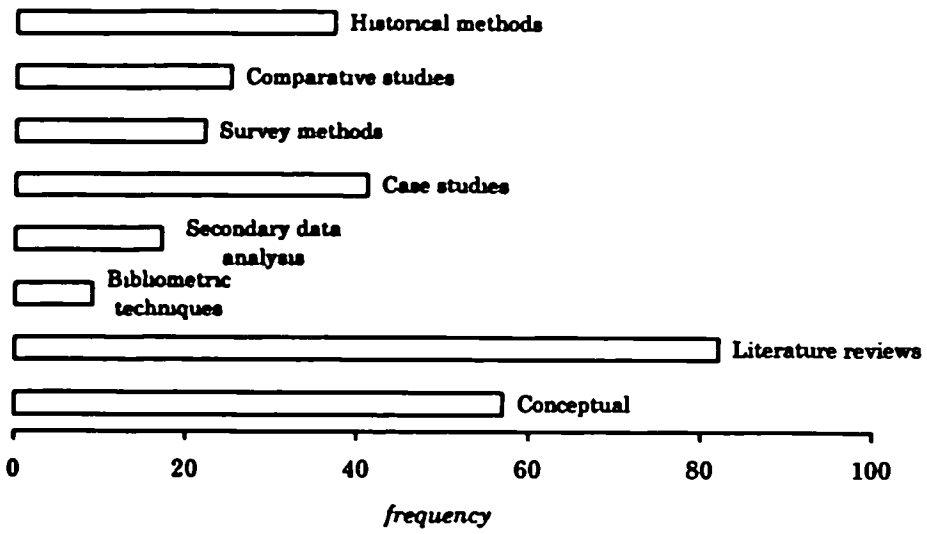


[74] Chi-square = 96.3, d.f. = 4, $p < 0.001$; Cramer's V = 0.35, $p < 0.001$

Following the approach adopted by Peritz (1980), a distinction was made between papers which exhibit "systematic method with the purpose of eliciting new facts, concepts or ideas" (i.e. those that are research-based) and those where no systematic method is evident (opinion papers). Figure 9.5 above shows that the bibliography is more or less equally divided into the two kinds of paper[75].

The most frequent research methods represented in the document test collection (Figure 9.6) are reviews of the literature (28.3 per cent), conceptual research strategies (19.7 per cent), case studies (14.1 per cent) and historical methods (12.8 per cent):

Figure 9.6: Research method (*n* = 290)



In an earlier stage of this doctoral research, Rowlands identified and described five broad methodological strands that were evident from a review of information policy publications and suggested that 'issues and options' was the most common research strategy (Rowlands, 1996).

[75] This finding is almost identical with Holowaty's analysis of 39 core library and information science journals (published in 1995) which found that 54 per cent were research articles and 46 per cent non-research (Holowaty, 1996).

Figure 9.7: Information policy research strategies (n = 284)

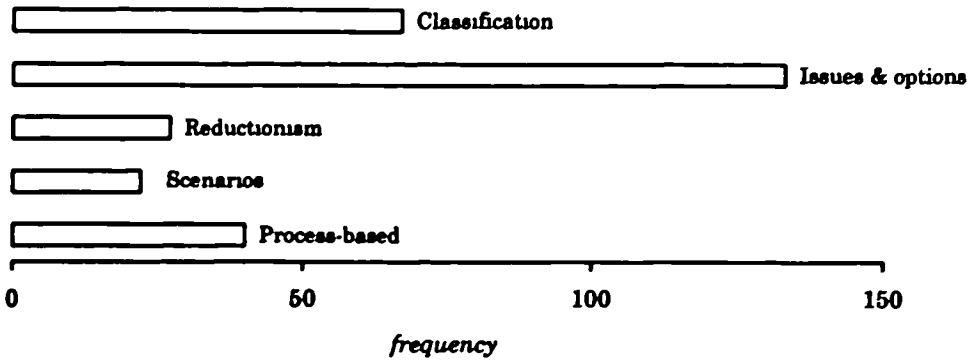
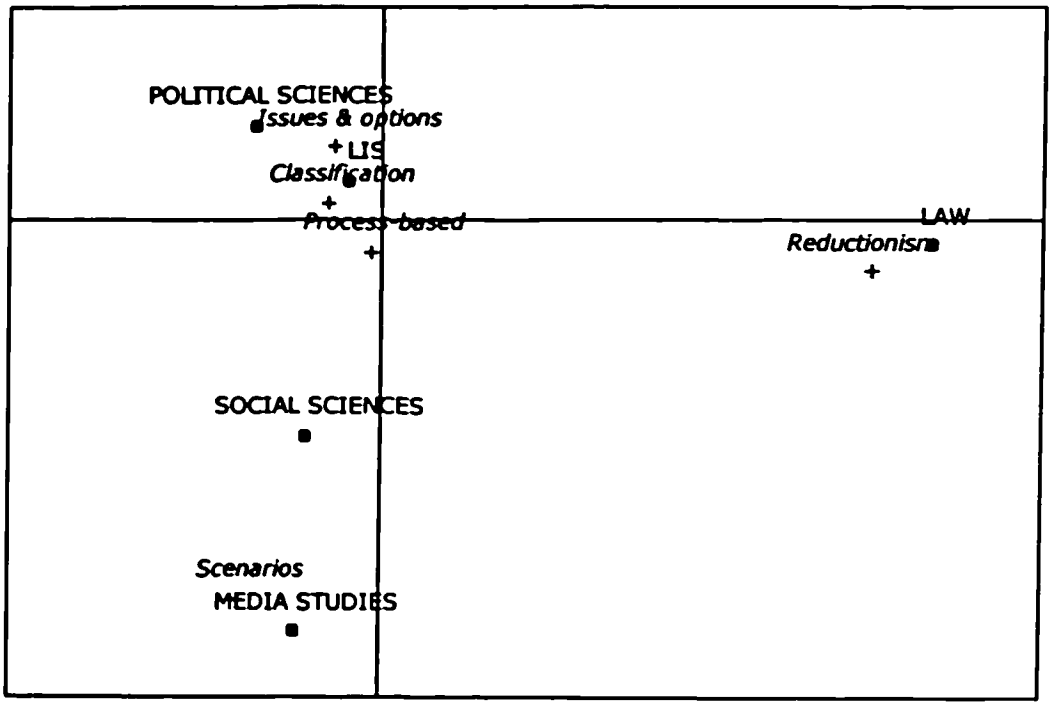


Figure 9.7 confirms this intuition, at least in relation to the research papers represented in the experimental bibliography.

Figure 9.8: ANACOR: research strategy by journal category[76]



The pattern of distribution of Rowlands' research strategies by journal category (Figure 9.8) opens up some interesting lines of speculation. Despite the values for chi-squared obtained[77], it is not possible to accept or reject the null hypothesis that the two variables are independent with full confidence, since more than half

[76] This two-dimensional mapping accounts for 97.8% of the total variance on two factorial axes, x (75.9%) and y (21.9%).
[77] Chi-square = 51.2, d.f. = 16, $p < 0.001$; Cramer's V = 0.38, $p < 0.001$

the cells (56 per cent) have an expected frequency of less than five. However, for what it is worth, the correspondence analysis may suggest that research papers in LIS journals are more likely to be grounded in classification, case studies and 'issues and options' than are the papers in other literatures. There is also, perhaps, an indication that the legal information policy literature may be quite strongly differentiated from the LIS journals in terms of the kinds of research strategies employed.

9.4 Aims, scope and intent of research

This section explores the stated aims, scope and intention of the research-based papers (only) in the document test collection, drawing on some of the conceptual frameworks presented in Chapter 4. The objective is to characterise the literature not by content, or research method, but by the analytical style and perspectives adopted by the authors. A subsidiary objective is to provide empirical evidence to substantiate or refute various claims that are commonly asserted in the literature—such as the purportedly highly national focus of much information policy research (Hill, 1995); the mis-alignment between research and the practical needs of policy-makers (Trauth, 1986; Strachan & Rowlands, 1997); and the paucity of research which evaluates the outcomes of information policy (Burger, 1993; Rowlands, 1997).

Defining the scope and extent of information policy is a widely acknowledged problem. Kristiansson (1996), like many other writers, draws a fundamental distinction between information policy issues according to the scale or level at which these issues should most appropriately be handled ranging from the global to the regional or sectoral:

Figure 9.9: Information policy scale (n = 771)

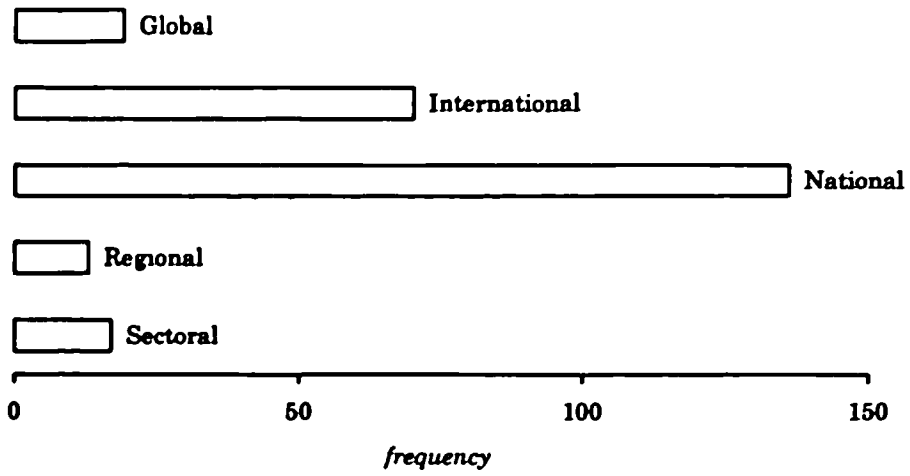
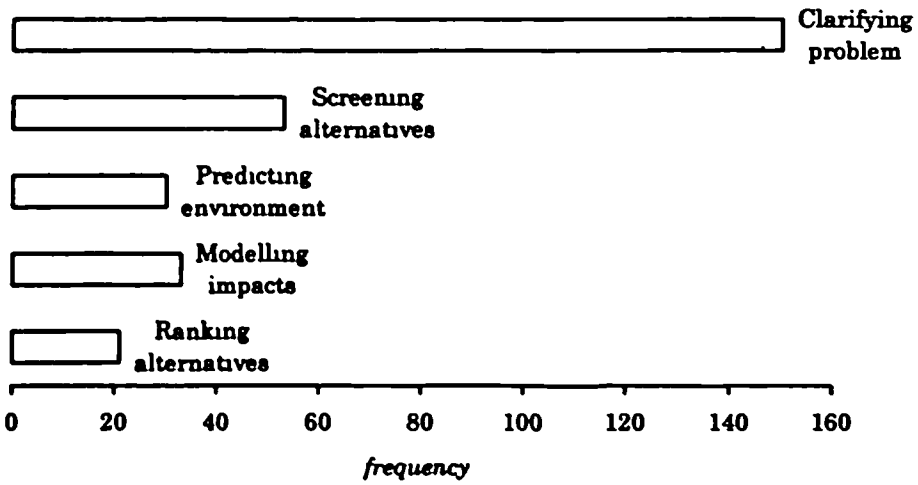


Figure 9.9 shows that the most frequent focus of articles in the test collection is on issues and policies at the national level. Relatively few papers engage with issues or events at either extreme of the spectrum: globally or sub-nationally. This appears to be a structural feature of the information policy test collection—no significant change in the distribution of articles by scale is evident over time, by author region or broad subject heading.

Figure 9.10: Main objective of article (n = 287)

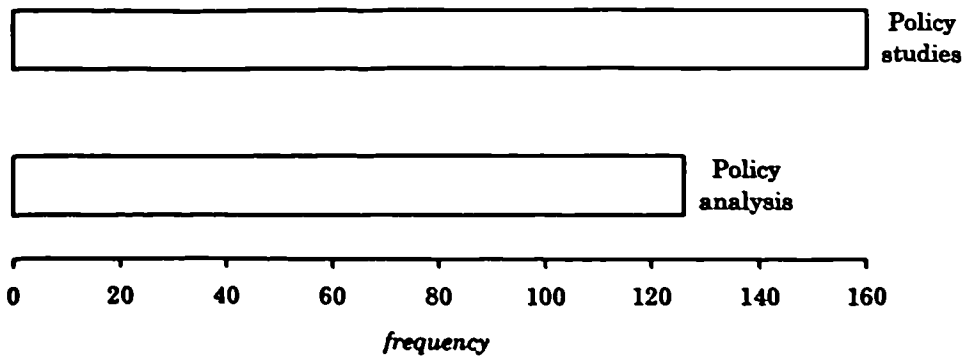


If much information policy research is oriented to national concerns and issues, can it also be that it is aligned to the needs of politicians and other policy-makers? A recent and much-debated theme in the general policy science literature focuses on the interface between the 'two communities' of policy

researchers and policy-makers. Quade (1976) has developed a five-point typology which locates the role of the analyst in relation to the needs of policy-makers: this forms the basis for the data presented in Figure 9.10 on the previous page. In Quade's scheme, a spectrum of engagement with the real world of policy-making is implied: 'Clarifying the problem' being the mode of analysis which is closest to the world of academic research, 'Ranking policy alternatives' closest to the immediate needs of policy-makers. The findings presented above suggest that the perspective adopted in the majority of research articles is one of seeking clarification and understanding rather than offering specific policy proposals and recommendations. This is not to imply, of course, that such work has no influence on policy-making, merely that any effect is likely to be conceptual rather than instrumental.

Gordon, Lewis & Young (1977) offer an alternative framework for understanding the underlying motivations of policy researchers. This draws a fundamental distinction between *policy studies* and *policy analysis*. Policy studies are studies *of* policy; they are motivated by curiosity rather than any explicit intention to shape the course of events. Policy studies embrace such activities as understanding the factors which determine policy, analyses of policy content, and policy monitoring and evaluation activities. In contrast, policy analysis actively seeks to influence the policy agenda; it is research *for* policy and is an integral part of the lobbying and influencing process.

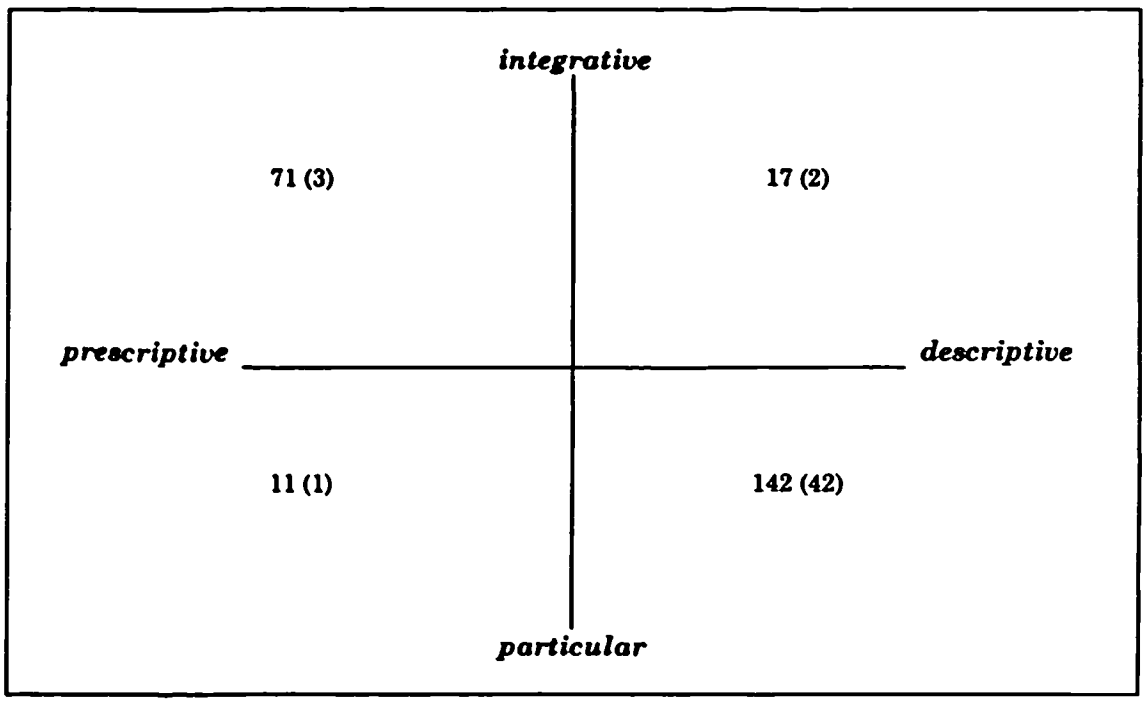
Figure 9.11: Analytical focus (n = 286)



The results presented in Figure 9.11 show that the articles in the document test collection tends towards the policy studies approach (55.9 per cent) rather than policy analysis (44.1 per cent). It is doubtful whether this finding could be generalised to the information policy research community at large; by its very nature, much policy analysis is proprietary, confidential and highly time-sensitive. Publication may be to a very restricted audience and take the form of informal materials rather than the serials literature.

One of the very few authors to tackle the nature and motivations behind information policy research is Trauth (1986). Trauth developed a highly generalised description of the information policy literature, locating studies in a two-dimensional matrix according to their scope (vertical axis) and intent (horizontal axis).

Figure 9.12: Scope and intent: changes since 1985 (n = 289)

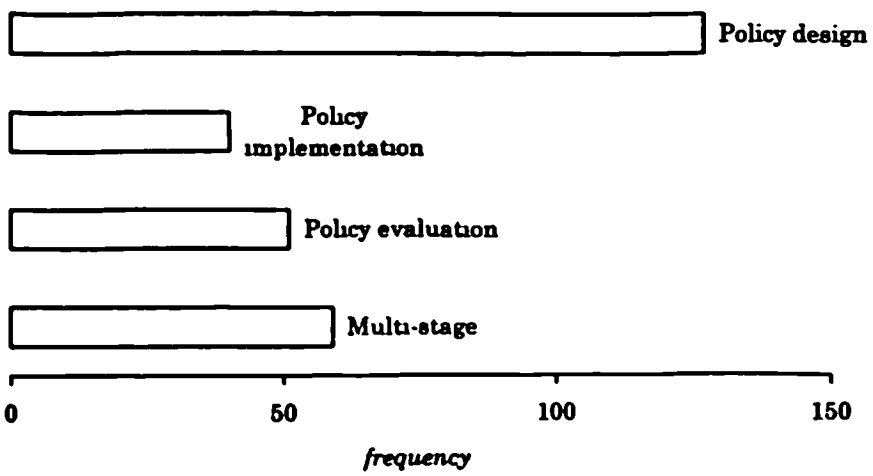


Trauth concluded in 1986 that most published information policy research was located in the bottom right quadrant of her model (i.e. particular and descriptive). She argued that there was a gap in the top left-hand quadrant; for studies which were both highly integrative and focused on the immediate needs of policy-makers. Test collection research papers are allocated to each of

Trauth's quadrants in Figure 9.12; in each case, the first number indicates the frequency of papers published in 1986 forward, while the number in brackets shows those papers published before 1986. The findings offer empirical evidence which tends to support Trauth's original claims; certainly the majority of studies at this time were located bottom right and, in the period since 1985, there does indeed appear to have been a substantial shift in favour of the top right-hand quadrant as she had hoped.

Another important dimension of information policy research is its temporal aspect; the stage of the policy cycle which is under investigation. The majority of research papers in the test collection appear to focus on the earlier stages of policy planning and design, rather than on the subsequent implementation or evaluation of existing policies (Figure 9.13).

Figure 9.13: Stage in policy cycle (n = 277)



Interest in policy design issues appears to be intensifying (Table 9.5 overleaf), possibly as a direct result of developments in the information infrastructure and the recent emergence of explicit Information Society policies (Oppenheim, 1996; Moore, 1997).

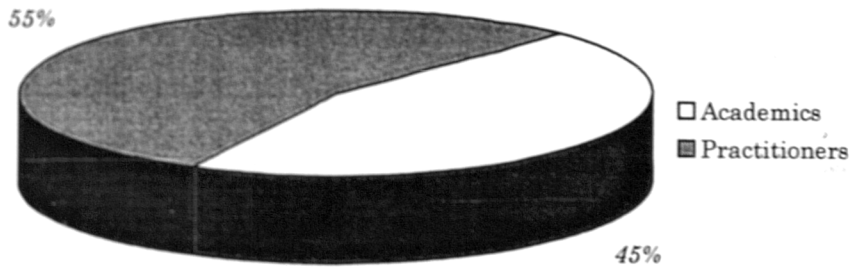
Table 9.5: Cross-tabulation: policy stage by time span[78]

	Design	Implementation	Evaluation	Multi-stage
Before 1986	11	12	14	10
	-3.4	*2.4*	*2.2*	*0.0*
1986 forward	116	28	37	49
	3.4	*-2.4*	*-2.2*	*0.0*

9.5 Patterns of authorship

Information policy research and scholarship is by no means the exclusive preserve of academics in traditional university settings. Figure 9.14 shows the distribution of first-named authors according to whether their corporate address is an institute of higher education ('academic') or another type of institution ('practitioner'):

Figure 9.14: Author status (n = 747)



Not surprisingly, a rather different pattern emerges if research-based papers only are considered: academics are responsible for 60 per cent of these, although it should be pointed out that non-academics still make a significant contribution to the bibliography, with a total of 114 research papers.

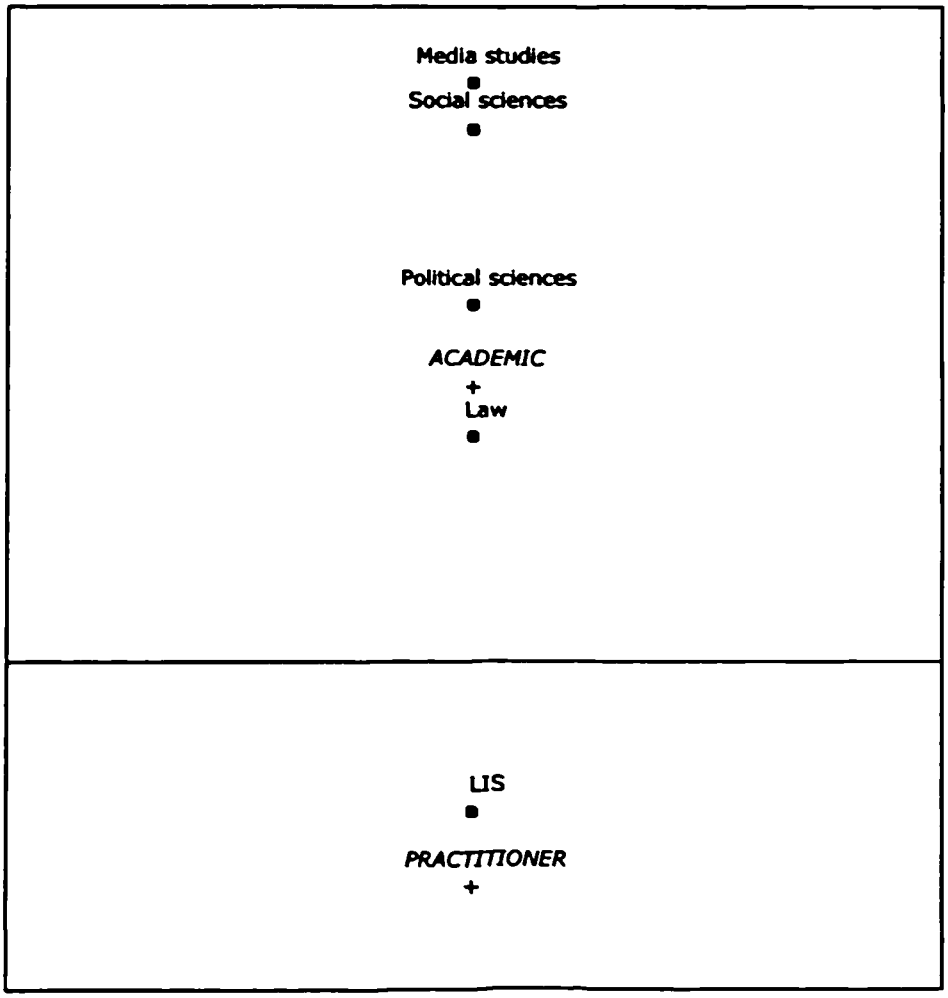
Author status shows an interesting pattern of distribution across *ISI* journal categories, as can be seen in Table 9.6 and Figure 9.15 on the next page:

[78] Chi-square = 16.1, d.f. = 3, $p < 0.005$; Cramer's V = 0.24, $p < 0.005$

Table 9.6: Cross-tabulation: author status by journal category[79]

	Academics	Practitioners
LIS	207	326
	-5.4	*5.4*
Media studies	18	8
	2.5	*-2.5*
Law	43	36
	1.8	*-1.8*
Political sciences	36	24
	2.4	*-2.4*
Social sciences	33	16
	3.2	*-3.2*

Figure 9.15: ANACOR: author status by journal category



[79] Chi-square = 32.9, d.f. = 6, $p < 0.001$; Cramer's V = 0.21, $p < 0.001$

Information policy authors publishing in LIS journal titles are much more likely to be based outside universities than authors publishing in any of the other journal categories. Media studies and social science journals are the most unlikely to include articles by practitioners.

Another interesting finding is presented below as Table 9.7. This shows that the balance of academics to practitioners has altered markedly over time, with a significant recent shift in favour of academic authorship. It is noted here that age of publication was also found to be an important discriminant factor in predicting Ward cluster membership in Chapter 7 (see Table 7.7, p.138).

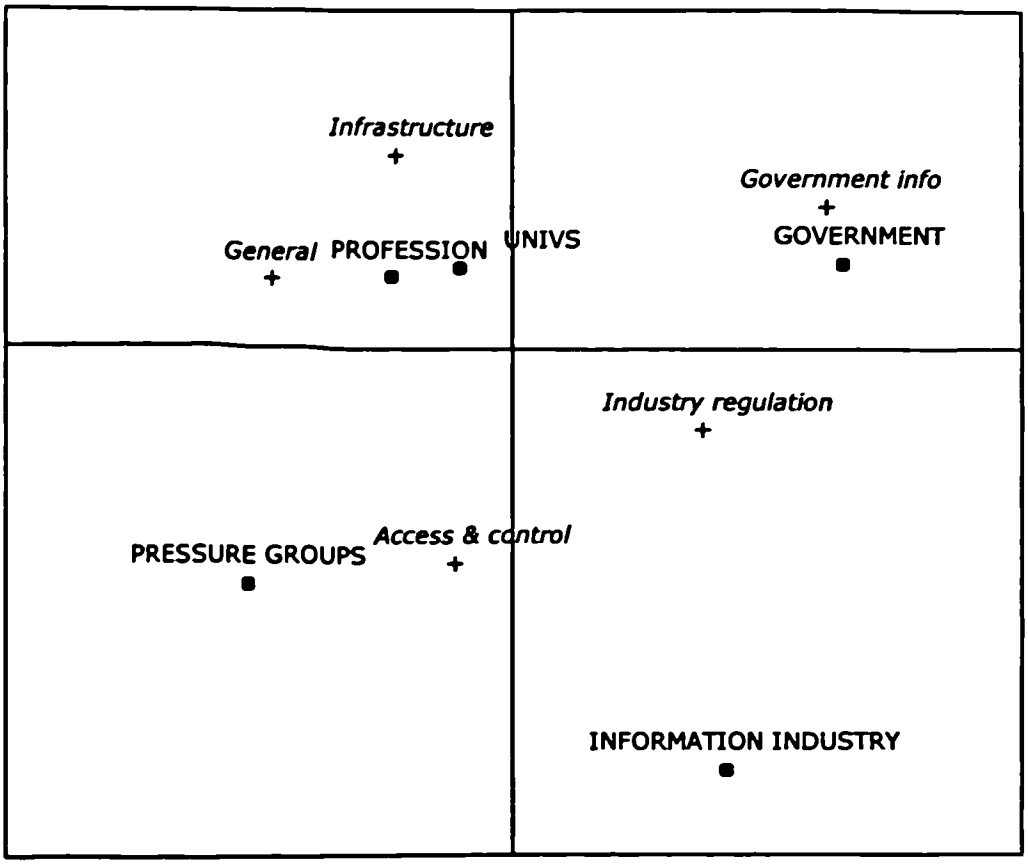
Table 9.7: Author status by time span[80]

	Before 1986	1986 forward
Academics	72	265
	-3.4	*3.4*
Practitioners	265	276
	3.4	*-3.4*

The final results presented in this Chapter indicate a significant and meaningful relationship between corporate author status and broad subject heading (Figure 9.16 overleaf).

[80] Chi-square = 14.3, d.f. = 1, $p < 0.01$; Cramers V = 0.14, $p < 0.01$

Figure 9.16: ANACOR: corporate author status by broad subject[81,82]



The subject profiles of the university authors (dominated by LIS schools) and library professionals are remarkably similar, both to one another and to the average for all corporate authors. Authors in these two sectors tend to be associated with general articles and papers on information infrastructure issues. Papers by authors located in the information industry exhibit the most atypical subject profile, with a propensity to write on information access and control issues (standard residual, $r = 2.4$) and regulatory matters ($r = 1.8$). Unsurprisingly, authors in government agencies tend to be associated with papers on government information issues ($r = 3.7$) and pressure groups with information access and control ($r = 2.3$).

[81] Chi-square = 49.5, d.f. = 16, $p < 0.01$; Cramers V = 0.13, $p < 0.01$

[82] This two-dimensional mapping accounts for 89.6% of the total variance on two factorial axes, x (54.4%) and y (35.1%).

Part Three
Synthesis of findings

Chapter 10: Conclusions

"In research the horizon recedes as we advance"[83]

10.1 Limitations of the study

The research findings arising from this study are inevitably provisional and require careful qualification. The limitations, some more serious than others, derive both from the nature of the test collection and the experimental design. The most obvious limitation of the work presented here stems from the decision to build the test collection from a single data source—the *Social Science Citation Index*. This decision was justified earlier (p.74), largely on the grounds of practicality and expediency. However, as already noted, one of the recurrent criticisms of bibliometric studies based on *ISI* data is that they are influenced by opaque editorial policies and decision-making processes internal to *ISI*. This raises issues which have been widely rehearsed in the bibliometric literature: there is certainly evidence, for example, that *ISI* citation indexes are biased in favour of the Anglo-American research literature (see, for example, Kärki, 1997). In a study which covers such an extended time period as the current one (1972-1996), doubts must also be raised about the stability (or, more likely, the instability) of *ISI* editorial policy over time. This source of uncertainty, which has not been investigated, raises serious concerns about the validity of any conclusions based on longitudinal analysis of the test collection.

Another factor which may have distorted the test collection is the responsibility of the author rather than *ISI* editorial staff. This concerns the effectiveness of the search strategies and relevance judgements used to identify and screen articles for inclusion in the experimental bibliography. The literature review identified concerns in the information policy research community regarding the field's poorly developed epistemological foundations and its sometimes confusing terminology (see Browne, 1997a, for an exposition of the problem). This,

[83] Mark Pattison, *Isaac Casaubon* (1875), Chapter 10.

together with the highly pervasive nature of information policies (Braman, 1990), means that the subject nomenclature from which search terms were extracted may be an idiosyncratic invention of the author. It is hoped that the test collection is at least broadly representative of the field of information policy, but this issue is unresolved and hardly trivial. The very limited bibliometric evidence available (see research finding 1 below), suggests that the test collection may be reasonably comprehensive, but this, even if it were to be confirmed, scarcely addresses the problem noted above. Coleman (1993) argues that from a document retrieval point of view, bibliometric studies of what he calls 'word specialties' are acutely problematic, especially in terms of the levels of recall that can reasonably be expected. These issues are compounded by *ISI's* understandable reluctance to provide a common indexing platform or thesaurus tools—*SSCI* records only contain index terms and descriptors where these are provided by the author. No system of broad subject headings is applied, with the exception of proprietary journal categories—however, the unit of analysis here is the whole journal, not the individual articles which comprise them, a common source of confusion in bibliometric studies. The limitations of the author cocitation study were aired in the conclusions to Chapter 8. The main issue here stemmed from the relatively low rates of cocitation encountered in the test collection as compared with the literatures typically selected for published investigation. Another problem with the author cocitation methodology, and one which limits its utility as a tool in science policy, stems from its reliance on highly cocited authors—who, almost by definition are already well-established, and this tends to present a conservative, backward-looking picture of the field under investigation.

Webb and others (1966) discuss two fundamental weaknesses in quantitative social research which are relevant to the work presented here: problems of internal and external validity. Both are a function of uncertainty. A study with high internal validity is one that would be able to explain observed differences within a sample with ease—the associated level of uncertainty would be low since most factors influencing the study had been fully taken care of. Naturally, studies of this kind are rare and most research wrestles as best it can with a degree of uncertainty. At many points in this thesis, apparently significant

bibliometric findings have been presented for which it has proved almost impossible to find a convincing and adequate interpretation. Wherever possible throughout the text, effort has been made to try to contextualise the bibliometric findings by reference to studies of other literatures. These comparisons should be read with considerable caution: King (1987) notes that there is a lack of commonly agreed guidelines or accepted standards for bibliometric research and argues, therefore, that there are few, if any, field-independent indicators in the literature.

The other set of problems in quantitative social research relates to poor external validity—an inability to generalise any observed differences within the sample to other samples. It has been difficult to resist the temptation to draw conclusions from this study without at least implying that they may be generalisable to the information policy research community as a whole. A major constraint on making any such claims arises from the study's self-imposed exclusive focus on the serials literature—to the author's knowledge, no specific work has been done on the information-seeking or communication behaviour of the information policy research community. Such work, were it to be done, would provide an essential context for this study in terms of the significance attached to the serials literature as a mechanism for scholarly communication. Monographs were not considered in this thesis, neither were other less formal modes of communication such as dissertations and theses, commissioned reports, submissions to legislative bodies, articles in less prestigious, non-refereed journals and so on. It is therefore dubious that the findings of this thesis can be generalised to the serials literature as a whole, let alone to the international community of information policy scholars.

10.2 Summary research findings

This section brings together the principal research findings from Chapters 6-9 in summary form:

1—The test collection is reasonably comprehensive[84]

The bibliography appears to be reasonably comprehensive—at least in so far as the journal productivity bibliograph (Fig.6.16, p.104) fails to show any indication of a Groos droop.

2—Documentary scatter is relatively low

Compared with published studies of other literatures, the information policy test collection appears to be relatively highly clustered, as measured by Brookes' clustering index (Ch.6, footnote 32, p.108) or by Coleman's article/journal density ratio (Tab.6.14, p.123). In other words, the test collection shows a relatively low degree of documentary scatter, rather less than might have been anticipated of a loosely-defined 'word specialty'.

---

[84] Incomplete bibliographies may be expected, from first principles, to show a deviation from linearity—the so-called Groos or terminal droop effect. At such a point, the bibliograph begins to flatten out as the number of new articles yielded falls. This phenomenon is usually interpreted as being due to the bibliographer missing relevant journal titles (in fact, missing productive non-singleton journals). No such effect can be seen in the experimental bibliograph (see Fig.6.16, p. 104) but this, by itself, is hardly conclusive proof that the bibliography is relatively complete.

The Bibliometrics Toolbox contains a routine for estimating the theoretical size of a literature—subject bibliographies are rarely, if ever, complete—based on a mathematical model developed by Egghe (1990). Egghe claims that it is possible to use this model even in cases where a Groos droop is not evident:

	Egghe's estimate of the theoretical size of the literature	*Rowlands' information policy bibliography*	*Difference*
Articles	846	771	75
Serial titles	249	181	68

These results suggest that the information policy bibliography is reasonably comprehensive: the differences between Egghe's predictions for the number of articles and serial titles are relatively small. More significantly, the ratio of 'missing' articles to 'missing' journals (1.10) suggests that any incompleteness in the bibliography is due to a failure to recover singleton rather than highly productive journal titles. This may explain why no terminal droop was observed in the bibliograph.

3—Signs of natural structure in the test collection

The results of the hierarchical cluster analysis (Ch. 7) strongly suggest that the test collection is characterised by an underlying, non-random, structure (it is possible to predict Ward cluster membership with 87 per cent probability on the basis of five discriminant factors: immediacy; nationality; journal category; author productivity; and age of article—Tab.7.4, p.134). The test collection is thus probably best regarded as a series of up to six (possibly incomplete) joint bibliographies sharing common journals but not common papers.

4—The test collection has a strongly anglophone bias

The bibliography comprises articles almost exclusively in the English-language (Fig.6.2, p.88) and production is dominated by US authors (Fig.6.3, p.89) and institutions (Fig.6.19, p.117).

5—Information policy is a hot growth topic

The pattern of growth in the test collection (Fig.6.7, p.92) suggests that the information policy serials literature is doubling in volume roughly every six years. Article production in the test collection exhibits a faster rate of growth than the general social sciences literature from which it was drawn (Fig.6.8, p.92). The North American component of the test collection has grown at a consistently faster rate than any other regional grouping (Fig.6.4, p.90).

6—Information policy is most heavily represented in LIS journals

The test collection articles are drawn predominantly from library and information science journal titles (Fig.6.5, p.90). The journals in the core Bradford zones are mainly LIS titles, while the outer ('scatter') zones tend to draw more heavily on non-LIS titles (Tab.6.5, p.110) However, none of the Bradford core journals were judged to be specific to the field of information policy.

7—Broad topics are unevenly distributed across the test collection

The distribution of information policy topics across the test collection was found to be highly assymetric in the content analysis study. Broad subject headings are strongly associated with author region (Tab.9.1, p.191; Fig.9.3, p.192) and with *ISI* journal category—the subject profiles of library and information science journals being strongly differentiated from titles in law and the social sciences (Tab.9.2, p.194; Fig.9.4, p.195). These findings parallel those of the hierarchical cluster analysis (Tab.7.6, p.137) and the author cocitation studies (Ch.8). The subject profile of the test collection has altered significantly since 1985, shifting from an emphasis on information access and control issues to government information, information infrastructure, and regulatory issues (Tab.9.4, p.197).

8—Academics / library professionals adopt the widest viewpoint

Of all author groups, academics and library professionals exhibit the broadest-based subject profiles. These are strongly differentiated from writers attached to government, the information industry or pressure groups, whose interests are more specialised (Fig.9.16, p.208).

9—Practitioners play a major role in information policy scholarship

The majority of articles in the test collection are written by practitioners rather than by academics (Fig.9.14, p.205). Practitioners contribute a sizeable number of research papers, although, unsurprisingly, fewer than academic authors. The overall pattern of author productivity seems to occupy an intermediate position between what is held to be typical of the natural and social sciences (Tab.6.8, p.113).

10—Information policy is not a unified field of study

The principal components analysis of the author cocitation data (Tab.8.6, p.164) identified four hidden factors (interpreted as 'Government Information'; 'Scientific & Technical Information'; the 'Social Implications of ICTs'; and 'Information Infrastructure & Regulation'). This typology was broadly supported by the outcome of the hierarchical cluster analysis (Fig.8.8, p.168). An important feature of the cocitation data is the low degree of intercorrelation between the four factors which, it is concluded, are effectively independent (Tab.8.7, p.166).

11—Journal categories may be differentiated by research strategy

The test collection is more or less equally split between research-based and opinion articles (Fig.9.5, p.197). Literature reviews, conceptual strategies, case studies, historical methods and comparative studies are the predominant research methods represented (Fig.9.6, p.198). Some differentiation between *ISI* journal categories was found in relation to the distribution of information policy research strategies (using Rowlands' typology): 'issues and options' being particularly associated with the library and information science and political science literatures, while law appears to be more closely associated with reductionist strategies (Fig.9.7, p.199).

12—The primary focus of most articles is on national concerns

The primary focus of most articles in the test collection is on issues and events at national level; a finding which appears to be independent of time, geography or broad topic (Fig.9.9, p.201). Most articles tend to deal primarily with policy design rather than its implementation or evaluation (Fig.9.13, p.204) and there is some evidence of a trend in this direction (Tab.9.5, p.205).

13—A gradual shift from policy studies to policy analysis

Overall, the test collection emphasises articles written from a policy studies rather than a policy analysis viewpoint (Fig.9.11, p.202). There is, however, some evidence that a gradual shift in perspective has occurred since 1985—from studies which are essentially descriptive to those which advocate policy solutions, and from research which is narrowly disciplinary-bound to more broadly-based integrative studies (Fig.9.12, p.203).

14—Author cocitations are highly patterned

The multidimensional scaling (ALSCAL) representation of the author cocitation data (Fig.8.16, p.185) was judged to be successful—in terms of the associated statistical indicators, the ease with which a *prima facie* interpretation of the data could be developed, and the corroborating evidence of the validation questionnaire. The author cocitation study suggests the existence of two major information policy schools, differentiated by geography and broad topic profile.

15—Author cocitation data reveals social and cognitive structure

The validation questionnaire administered during the author cocitation study revealed highly significant relationships between the Euclidean distances used to represent author dissimilarity in the ALSCAL mappings and (a) the degree of mutual author recognition (Fig.8.12, p.181), and (b) the intensity of the social and collaborative ties between authors (Fig.8.13, p.182). The questionnaire generally confirmed the allocation of authors to specific topic clusters (Tab.8.10, p.183) and suggests that it is reasonably valid to interpret the data either from a social or a cognitive viewpoint.

16—Sole authorship is the norm in information policy

The test collection is characterised by a very high incidence of sole authorship (Tab.6.9, p.115), a feature which it shares with the library and information science literature (Ch.6: footnotes 34,35). The incidence of jointly-authored papers is low in comparison with the general social science literature (p.115), however, and this appears to be a structural feature of the information policy test collection—there is no significant evidence of any conformity with a general trend in the social sciences for the proportion of jointly-authored papers to increase over time (Tab.6.10, p.115).

17—The information policy research base is highly dispersed

Information policy research is very thinly dispersed in institutional terms, with 341 unique corporate addresses identified for first-named authors. Government institutions form the most concentrated research sector, while academia is the most highly dispersed (Tab.6.12, p.121). Universities comprise the largest single institutional grouping, followed by government, the information industry, library professionals and pressure groups (Tab.6.12, p.121). Although the largest grouping, university-based authors contributed less than half of the test collection (Fig.6.21, p.119). Surprisingly, the institutional profile of the test collection has become more even more highly dispersed in the period since 1985 (Tab.6.12, p.121).

18—Research effort is most highly concentrated in the USA

Most of the top-ranking corporate authors in terms of article production are US institutions (Tab.6.11, p.120). The UK has the largest number of European corporate authors, followed by Germany and France (Fig.6.20, p.118). A survey by Stroetmann (1992) suggests that European information research sector may be showing signs of incipient concentration.

19—Little published evidence of collaborative research activity

Where it occurs, most published examples of research collaboration take place between authors located in the same institution; collaboration between institutions or between authors working in different sectors is relatively uncommon (Fig.6.22, p.122). This pattern has not changed significantly over time.

20—Rapid obsolescence of information policy articles

The citation half-life of a sample of information policy articles published between 1981-85 was found to be only three years (Fig.6.15, p.101), a surprisingly low value when compared with published estimates for a wide range of other literatures (Tab.6.13, p.123).

21—Citation practices in information policy are highly immediate

The information policy test collection exhibits unexpectedly high values for Price's index in comparison with fields such as science policy, scientometrics or information science (Fig.6.13, p.97) and are more typical of values encountered in the hard sciences. Moreover, the high values for Price's index do not appear to be a function of the small size of the information policy archive (Fig.6.14, p.98). Cozzens (1985) argues that high levels of immediacy may be interpreted as a function of 'intellectual focus and excitement'.

22—The profile of information policy is becoming more 'scholarly'

The test collection may be characterised as being highly 'scholarly' as defined by its Windsor ratio (Fig.6.9, p.94). The value of the Windsor ratio has increased since 1985 (Fig.6.10, p.94) and research-based papers are slowly displacing opinion papers (Fig.6.11, p.95; Tab.9.7, p.207) in the literature. Authors writing in library and information science journal titles are very much more likely to be non-academics than in any other *ISI* journal category (Tab.9.6, p.206; Fig.9.15, p.206).

10.3 Discussion of research findings

In bibliometric terms, it is concluded that the information policy test collection appears to be reasonably comprehensive (research finding 1) and compact (research finding 2). Some of its characteristics are rather unusual and certainly atypical of other social science literatures—the most surprising findings here are those relating to citing behaviour: the rapid citation half-life (research finding 20) and the high values associated with Price's index (research finding 21). In Price's typology, information policy would be a good example of a highly immediate or 'ephemeral' literature, of a kind most commonly observed in the natural rather than the social sciences. It is not easy to interpret these findings. They may support a view that information policy is essentially a set of problem-solving, rather than academically-motivated, activities driven primarily by external social need. The fact that practitioners are heavily represented in the test collection (research finding 9) lends support to this view. Also, the finding that information policy is growing at a faster rate than the general social science literature (research finding 5) may be a response on the part of the research community to the high political profile which now surrounds information-related issues. An alternative, but related explanation, is the suggestion by Cozzens (1985) that high levels of immediacy may be encountered in disciplinary areas that are undergoing revolutionary, paradigmatic change.

One of the explicit research aims of this research was to address the question 'To what extent is information policy a distinct specialty in its own right, or does it comprise several specialist areas that are primarily dependent on other disciplines?'. The empirical evidence presented here shows that information policy is represented across a wide range of social science literatures: including library and information science, law, politics, public administration, communications studies and business studies. The profiles of these literatures are strongly differentiated, however, in respect of the topics covered, the professional status of the authors, and the research strategies employed (research findings 3, 7, 11).

One sign of the maturity of a field is the existence of its own specialised journals. Although there is no general purpose 'Journal of Information Policy', there are a number of titles with a strong, if sector-specific, information policy bias (notably *Government Information Quarterly* and *Journal of Government Information*). Information policy is very heavily represented in the library and information science literature, which also contributes disproportionately to the inner Bradford zones of journal productivity (research finding 6). These findings tend to support the view of Burger (1993), that library and information science is the natural disciplinary home for information policy. Further support for this view comes with the finding that the subject profile of articles in LIS journals is the least specialised of any of the literatures studied (research findings 7 and 8).

Even if one accepts that LIS is the home discipline for information policy, it is very doubtful from the evidence presented here that information policy could in any way be regarded as a unified field of study. While the results of the author cocitation study suggest that information policy scholarship has real social and intellectual structure, the analysis reveals clusters of prominent authors which appear to be distinct and to exhibit little interaction (research finding 10). Faced with the evidence from the author cocitation study and the parallel findings of the cluster and content analyses, it is hard to escape the conclusion that information policy is chiefly concerned with national concerns (research finding 12) and that research and scholarship are primarily organised along geographical and traditional disciplinary lines.

A further aim of this research was to investigate the kind of institutional arrangements that support information policy research, since the degree of institutionalisation of a discipline or specialty is generally held to be one indicator of its maturity and status. Here the results are striking—information policy appears to be very highly dispersed in an institutional sense (341 unique corporate addresses were identified for first authors). Indeed the trend seems to be for the field to have become more rather than less institutionally dispersed over time—especially in the university sector. There are few signs of research concentration, particularly outside the USA, and little evidence from the published record of collaborative research activity—especially across disciplinary

or sectoral boundaries. The impression is one of scholars working in isolation. Sole authorship appears to be the norm (research finding 16) and, counter to the trend in the general social sciences literature, there is no evidence that collaborative authoring is on the increase. The meaning and implications of these findings are difficult to assess. It is certainly tempting to make broad generalisations about the field of information policy being weakly institutionalised, even ghettoised. However, it is also possible that the high levels of institutional dispersion encountered in the test collection are merely an artefact. It may simply be the case, given the problem-oriented nature of information policy, that some authors, who normally work on other aspects of public policy, information science or management, are temporarily drawn into the information policy realm in order to respond to a particular issue of the moment. It cannot possibly be assumed that all the authors represented in the document test collection necessarily identify themselves as information policy scholars!

The final research aim was to try to characterise information policy in terms of the model of knowledge production proposed by Gibbons and others (1994). In Gibbons' scheme, Mode 2 knowledge is generated across a wide range of institutional settings and not confined to the university (see research finding 17). Knowledge is produced in response to external social and economic need, rather than for its own sake according to traditional scientific and academic norms. Mode 2 knowledge production is characterised by a high degree of organisational diversity and by structures which are flexible and transient. Intense levels of collaboration, across institutions, sectors and disciplines are a further characteristic of Mode 2 knowledge production. From the very limited evidence in this thesis, information policy appears to share many of the characteristics that Gibbon and co-workers associate with Mode 2 knowledge production, although the low incidence of collaborative research and authorship should be noted. Clearly, more work is needed to fully explicate the organisational and sociological dynamics of information policy research. The findings of this study provide *prima facie* evidence to suggest, however, that Gibbons' scheme may offer a useful framework for organising bibliometric data at the levels of research fields and disciplines.

10.4 Directions for further research

In order to address some of the concerns expressed earlier over the limited internal and external validity of these findings, further work is needed. During the course of this research a number of avenues for possible further investigation opened up. The most pressing of these relates to the need for a wider context for the research findings. The notion that disciplines and specialties are, at least in part, social rather than intellectual constructions was touched upon in the literature review (Ch.2), and some evidence to support this view was provided in the form of the author cocitation study. This aspect requires separate investigation, perhaps along the lines of the semi-structured interview methodology developed by Becher (1990). Some of the areas that such a study might investigate include the backgrounds, careers, motivations and attitudes of information policy scholars, and the institutional arrangements, patterns of funding and reward systems which support their work. A qualitative investigation along these lines would provide a valuable context for the bibliometric evidence presented here.

On a more abstract plane, there several fundamental areas of theory where, currently, knowledge and understanding are poorly developed. Scholars in information policy are becoming increasingly self-reflexive, questioning the assumptions upon which their work is based and arguing the need for more robust value- and paradigm-critical approaches (Rowlands, 1996; Browne, 1997a,b). However, most writers acknowledge that the epistemological foundations of information policy are poorly understood.

Further progress in understanding the relationship between knowledge structures and knowledge communities is needed. A recent programmatic article by Hjørland & Albrechtsen (1995) opens up some interesting vistas here in proposing the concept of 'domain analysis'—a more holistic approach to understanding different areas of knowledge as thought or discourse communities. It would be especially interesting to try to locate bibliometric methods within this new approach.

There are unfortunately no studies in the literature which deal specifically with the information-seeking and communication behaviour of information policy scholars. Work of this kind is needed to contextualise the findings presented here, especially in relation to the role which the serials literature plays in information policy research and scholarship. The results of a study of this kind might indicate the respective functions of monographs, non-refereed journals, commissioned research reports, and other grey literature in the scholarly communication process.

Further work is also needed to extend the bibliometric findings presented in this thesis. For example, the absence of an author citation analysis is a major gap which needs to be filled. More could be done to extend the author cocitation study: possibly testing the stability of the model by adding new or removing existing new authors. An analysis of the networks of cocitations between core journals might offer further insights into the structure of the field, and make interesting comparison with the author cocitation study. Similarly, work on coauthorship (rather than author cocitation) might prove valuable, especially if the terms of engagement were sufficiently broad to include genealogies of information policy academics and their doctoral students, hence tracing the diffusion of ideas and perspectives through the scholarly community.

References

Aldenderfer, M. & Blashfield, R. (1984). *Cluster Analysis* (Beverley Hills, CA: Sage).

Aldhouse, F. (1997). Implementing information policies: some data protection experience. In: Rowlands, I. (ed.), *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 114-127).

Allen, D. & Wilson, T. (1997). Information systems strategy formulation in higher education. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 178-190).

Anderson, J.E. (1990). *Public Policymaking: an Introduction* (Boston: Houghton Mifflin).

Bailey, T.A. & Dubes, R. (1982). Cluster validity profiles, *Pattern Recognition 15(2)* 61-83.

Bates, B.J. (1990). Information systems and society: potential impacts of alternative structures, *Telecommunications Policy 14(2)* 151-158.

Bayer, A.E.; Smart, J.C. & McLaughlin, G.W. (1990). Mapping the intellectual structure of a scientific subfield through author cocitations, *Journal of the American Society for Information Science 41(6)* 444-452.

Bearman, T.C. (1988). Uses of scientific, technical and societal information by policy makers, *Knowledge in Society 1(1)* 27-53.

Becher, A. (1990). *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines* (Milton Keynes: The Society for Research into Higher Education / Open University Press).

Becher, A. & Kogan, M. (1980). *Process and Structure in Higher Education* (London: Heinemann).

Bellardo, T. (1980). The use of co-citations to study science, *Library Research 2(3)* 231-237.

Beniger, J.R. (1988). Information and communication: the new convergence, *Communication Research 15(2)* 198-218.

Bennett, J.R. (1987). *Control of Information in the United States: An Annotated Bibliography* (Westport, CT: Meckler).

Berelson, B. (1960). *Graduate Education in the United States* (New York: McGraw-Hill).

van den Besselaar, P. & Leydesdorff, L. (1996). Mapping change in scientific specialities: a scientometric reconstruction of the development of artificial intelligence, *Journal of the American Society for Information Science 47(6)* 415-436.

Bezold, C. & Olson, R. (1986). *The Information Millenium: Alternative Futures: a Report for the Information Industry Association* (Washington: Information Industry Association).

Borgman, C.L. (1990). Editor's introduction. In: Borgman, C.L. (ed.) *Scholarly Communication and Bibliometrics* (London: Sage, 10-27).

Borgman, C.L. & Rice, R.E. (1992). The convergence of information science and communication: a bibliometric analysis, *Journal of the American Society for Information Science 43(6)* 397-411.

Bottle, R.T. & Efthimiadis, E.N. (1984). Library and information science literature: authorship and growth patterns, *Journal of Information Science 9(3)* 107-116.

Bottle, R.T. & Gong, Y.T. (1987). A bibliometric study on the ageing and content typology relationship of the biochemical literature, *Journal of Information Science 13(1)* 59-63.

Bradford, S.C. (1934). Sources of information on specific subjects, *Engineering 137* (3550) 85-86.

Braman, S. (1988). *Information Policy and the United States Supreme Court* (University of Minnesota, PhD thesis).

Braman, S. (1989). Defining information: an approach for policymakers, *Telecommunications Policy 13(3)* 233-242.

Braman, S. (1990). The unique characteristics of information policy and their US consequences, In: Blake, V. & Tjoumas, R. (eds) *Information Literacies for the Twenty-first Century* (Boston, MA: Hill, 1990, 47-77).

Brookes, T.A. (1990). Literature core zones adjusted by impact factors, *Journal of Information Science 16(1)* 51-57.

Browne, M. (1996). Personal communication, 14 November.

Browne, M. (1997a). The field of information policy: (1) fundamental concepts, *Journal of Information Science* 23(4) 261-275.

Browne, M. (1997b). The field of information policy: (2) re-defining the boundaries and methodologies, *Journal of Information Science* 23(5) 339-351.

Burger, R.H. (1988). *The Evaluation of Information Policy: a Case Study Using the SATCOM Report* (University of Illinois at Urbana-Champaign, PhD thesis).

Burger, R.H. (1993). *Information Policy: a Framework for Evaluation and Policy Research* (Norwood, NJ: Ablex).

Burke, M.E. (1994). Introduction to scenario forecasting, In: *Proceedings of the European Business Information Conference 1994,* Paris, 23-25 March 1994 (London: TFPL Publishing, 225-231).

Burkert, H. (1992). The legal framework of public sector information: recent legal policy developments in the EC, *Government Publications Review 19(5)* 483-496.

Bush, V. (1945). *Science, the Endless Frontier: a Report to the President on a Program for Postwar Scientific Research* (Washington: National Science Foundation).

Bushkin, A.A. & Yurow, J.H. (1981). The foundations of United States information policy, In: Gassmann, H.P. (ed.) *Proceedings of the High Level Conference on Information, Computer and Communications Policies for the 80's,* Paris, 6-8 October, 1980 (Amsterdam: North-Holland).

Cabinet Office (1984). *Making a Business of Information* (London: Information Technology Advisory Panel).

Chartrand, R.L. (1986). Legislating information policy, *Bulletin of the American Society for Information Science 12(5)* 10.

Chubin, D.E. (1976). The conceptualisation of scientific specialties, *Sociological Quarterly 17(3)* 448-476.

Cole, P.E. (1962). A new look at reference scattering, *Journal of Documentation 18(2)* 58-64.

Cole, S. (1983). The hierarchy of the sciences, *American Journal of Sociology 89(1)* 111-139.

Coleman, S.R. (1993). Bradford distributions of social science bibliographies varying in definitional homogeneity, *Scientometrics 27(1)* 75-91.

Compaigne, B.M. (1988). Evolving technology and blurring boundaries: new perspectives for evaluating media competition, In: Compaigne, B.M. (ed.) *Issues in New Information Technology* (Norwood, NJ: Ablex).

Cottrill, C.A., Rogers, E.M. & Mills, T. (1989). Cocitation analysis of the scientific literature of innovation research traditions, *Knowledge: Creation, Diffusion, Utilization 11(3)* 181-208.

Crane, D. (1972). *Invisible Colleges* (University of Chicago Press).

Cozzens, S.E. (1985). Using the archive: Derek Price's theory of differences among the sciences, *Scientometrics 7(3/6)* 431-441.

Cronin, B., Davenport, E. & Martinson, A. (1997). Women's studies: bibliometric and content analysis of the formative years, *Journal of Documentation 53(2)* 123-138.

Cronin, B. & Pearson, S. (1990). The export of ideas from information science, *Journal of Information Science 16(6)* 381-391.

Culnan, M.J. and others (1990). Intellectual structure of research in organizational behaviour: 1972-1984: a cocitation analysis, *Journal of the American Society for Information Science 41(6)* 453-458.

Cunningham, S.J. & Bocock, D. (1995). Obsolescence of computing literature, *Scientometrics 34(2)* 255-262.

Dahlin, T.C. (1990). *Books and Documents on Government Information Policy: a Selected Bibliography* (Moniticello: Vance Bibliographies).

Dawes, S.S. (1991). *A Theory of Interagency Information Sharing: Policy Coordination* (State University of New York at Albany, PhD thesis).

Dean, M.C. (1980). *A Quantitative Analysis of Theory Change in Experimental Psychology.* Paper presented at the fifth annual meeting of the Society for Social Studies of Science, Toronto, 17-19 October 1980.

Department of National Heritage (1993). *The Secretary of State's Speech to the Library Association Umbrella Conference on 9 July 1993* (London: Department of National Heritage Press Office).

Donohoe, J.C.(1983). *Understanding Scientific Literatures: A Bibliometric Approach* (Massachussetts Institute of Technology Press).

Downs, A. (1957). *An Economic Theory of Democracy* (New York: Harper & Row).

Drott, M.C. (1981). Bradford's Law: theory, empiricism and the gaps between, *Library Trends 30(1)* 41-52.

Dubes, R. & Jain, A.K. (1979). Validity studies in clustering methodologies, *Pattern Recognition 11(4)* 235-254.

Duff, A.S. (1995). The 'information society' as paradigm: a bibliometric enquiry, *Journal of Information Science 21(5)* 390-395.

Dye, T.R. & Ziegler, L.H. (1990). *The Irony of Democracy* (Monterey: Brookes/Cole,).

Dye, T.R. (1984). *Understanding Public Policy* (Englewood Cliffs: Prentice-Hall).

East, H. (1997). The limitations of 'free' access in academia: making commercial databases available. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 221-231).

Easton, D. (1965). *A Systems Analysis of Political Life* (New York: Wiley).

Edge, D. (1977). Why I am not a cocitationist, *Newsletter of the Society for Social Studies in Science 2* 13-19.

Egghe, L. (1989). *A Note on Different Bradford Multiplicators* (unpublished manuscript, 1989, cited in Brooks, 1990).

Egghe, L. & Rousseau, R. (1996). Averaging and globalising quotients of informetric and scientometric data, *Journal of Information Science 22(3)* 165-170.

Eisenschitz, T.S. (1997). Legal and regulatory frameworks for information policy. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 128-145).

Eom, S.B. (1996). Mapping the intellectual structure of research in decision-support systems through author cocitation analysis (1971-1993), *Decision Support Systems 16(4)* 315-338.

Fox, M.F. & Faver, C.A. (1982). The process of collaboration in scholarly research, *Scholarly Publishing 13(3)* 327-339.

Galvin, T.J. (1994). Rights in conflict: public policy in an information age, In: *New Worlds in Information and Documentation: Proceedings of the 46th FID Conference and Congress 1992* (The Hague: FID).

Garfield, E. (1964). *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information).

Garfield, E. (1979). *Citation Indexing: Its Theory and Application in Science, Technology and Humanities* (New York: Wiley).

Garfield, E. (1985). In tribute to Derek John deSolla Price: a citation analysis of Little Science, Big Science, *Scientometrics 7(5)* 487-503.

Garfield, E.; Malin, M.W. & Small, H. (1978). Citation data as science indicators, In: Elkana, Y. and others (eds), *Toward a Metric of Science: the Advent of Science Indicators* (New York: Wiley, 197-207).

Geertz, C. (1983). *Local Knowledge* (New York: Basic Books).

Gibbons, M. and others (1994). *The New Production of Knowledge* (London: Sage).

Glänzel, W. & Schoepflin, U. (1995). A bibliometric study on ageing and reception processes of scientific literature, *Journal of Information Science 21(1)* 37-53.

Goffman, W. & Warren, K.S. (1969). Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures, *Nature 221 (5187)* 1205-1207.

Gordon, I., Lewis, J. & Young, K. (1977). Perspectives on policy analysis, *Public Administration Bulletin 25(1)* 26-35.

Gray, J. (1988). *National Information Policies: Problems and Progress* (London: Mansell).

Gray, J. (1989). National information policy: myth or magic? *Alexandria 1(3)* 21-30.

Gray, R. (1993). *Towards a Framework for a National Information Policy: Discussion Paper for LISC(E) Meeting to be held at the Department of National Heritage on 4 November 1993* (London: Rosemary Gray Associates).

Greenstein, S.M. (1992). Invisible hands and visible advisors: an economic interpretation of standardization, *Journal of the American Society for Information Science 43(8)* 538-549.

Griffith, B.C. (ed.) (1980). *Key Papers in Information Science* (White Plains, NY: Knowledge Industry Publications, 415-439).

Griffith, B.C. (1990). Understanding science: studies of communication and information. In: Borgman, C.L. (ed.) *Scholarly Communication and Bibliometrics* (London: Sage, 31-45).

Hagstrom, W.O. (1968). *The Scientific Community* (New York: Basic Books).

Haiqi, Z. (1995). Basic literature of acupuncture in MEDLINE: a bibliometric analysis. *Libri 45(2)* 113-122.

Hammond, K.R. and others (1983). The use of scientific information in public policy making, *Technological Forecasting and Social Change 24(4)* 287-297.

Hawkins, D.T. and others (1992). Forces shaping the electronic publishing industry of the 1990s, *Electronic Networking 2(4)* 38-60.

Hernon, P. & Relyea, H.C. (1968). Information policy, In: A. Kent and H. Lacour (eds), *Encyclopedia of Library and Information Science* Volume 48 Supplement II (New York: Dekker).

Hicks, D. & Katz, J.S. (1996). Systemic bibliometric indicators for the knowledge-based economy. Paper presented at OECD workshop on *New Indicators for the Knowledge-Based Economy,* Paris, 19-21 June, 1996.

Hicks, D. & Katz, J.S. (1997). *Collaboration and Diversity in UK Research* (http://www.susx.ac.uk/spru/hickskatz/t_of_con.html, 20 June 1997).

Hill, M. W. (1995). *National Information Policies and Strategies: an Overview and Bibliographic Survey* (London: Bowker-Saur British Library Research series).

Hirst, P. H. (1974). *Knowledge in the Curriculum: A Collection of Philosophical Papers* (London: Routledge & Kegan Paul).

Hjørland, B. (1992). The concept of 'subject' in information science, *Journal of Documentation 48(2)* 5-37.

Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis, *Journal of the American Society for Information Science 46(6)* 400-425.

Hogwood, B.W. & Gunn, L.A. (1984). *Policy Analysis for the Real World* (Oxford University Press).

Holowaty, I.D. (1996). *Content Analysis of 1995 Core Library and Information Science Journals* (City University, MSc dissertation).

Irvine, J. & Martin, B.R. (1989). International comparisons of scientific performance revisited, *Scientometrics 15(5/6)* 369-392.

Jacobson, R.E. (1989). *An "Open" Approach to Information Policy Making: a Case Study of the Moore Universal Telephone Act* (Norwood, NJ: Ablex).

Järvelin, K. & Vakkari, P. (1990). Content analysis of research articles in library and information science, *Library and Information Science Research 12(4)* 395-421.

Järvelin, K. & Vakkari, P. (1993). The evolution of library and information science 1965-1985: a content analysis of journal articles, *Information Processing & Management 2(1)* 129-144.

Johansen, G. (1997). Information, knowledge and research, *Journal of Information Science 23(2)* 103-109.

Julien, E., Robinson, D. & Tinline, J. (1992). *Bibliography for the National Summit on Information Policy* (Ottawa: National Library of Canada).

Kajberg, L. (1996). A citation analysis of LIS serial literature published in Denmark 1957-1986, *Journal of Documentation 52(1)* 69-85.

Kajberg, L. & Kristiansson, M. (1996). An overview of the field of information policy, *International Forum on Information and Documentation 21(1)* 5-9.

Kärki, R. (1996). Searching for bridges between disciplines: an author co-citation analysis on the research into scholarly communication, *Journal of Information Science 22(5)* 323-334.

Karni, R. (1983). A methodological framework for formulating information policy, *Information and Management 6(5)* 269-280.

Kerlinger, F.N. (1973). *Foundations of Behavioral Research* (New York: Holt, Rinehart & Winston).

Keynes, J.M. (1936). *The General Theory of Employment, Interest and Money* (London: Macmillan).

Kiesler, S.B. & Hinds, P. (1993). Technology, information and social behaviour, In: *The Knowledge Economy: the Nature of Information in the 21st Century: Annual Review of the Institute for Information Studies 1993-94* (Nashville, TS: Institute for Information Studies, 117-135).

King, A.R. & Brownell, J. (1966). *The Curriculum and the Disciplines of Knowledge* (New York: John Wiley).

King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation, *Journal of Information Science 13(5)* 261-276.

Kovach, E.J. (1978). Country trends in scientific productivity, In: *ISI's Whos Publishing in Science* (Philadelphia: ISI Press, 33-40).

Krippendorff, K. (1980). *Content Analysis: An Introduction to its Methodology* (London: Sage Publications).

Kristiansson, M. (1996). A framework for information policy analysis based on changes in the global economic forces, *International Forum on Information and Documentation 21(1)* 5-9.

Kruskal, J.B. & Wish, M. (1978). *Multidimensional Scaling: Quantitative Applications in the Social Sciences* (London: Sage).

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions* (University of Chicago Press).

Lancaster, F.W. (1991). *Bibliometric Methods in Assessing Productivity and Impact of Research* (Bangalore: Sarada Ranganathan Endowment for Library Science).

Langridge, D. (1991). Classifying knowledge, In: Meadows, A.J. (ed.), *Knowledge and Communications: Essays on the Information Chain* (London: Library Association, 1-18).

Lasswell, H.D. (1930). *Psychopathology and Politics* (University of Chicago Press).

Lasswell, H.D. (1948). *Power and Personality* (New York: W.W. Norton).

Lasswell, H.D. (1970). The emerging conception of the policy sciences, *Policy Sciences 1(1)* 3-14.

Latham, E. (1965). *The Group Basis of Politics* (New York: Octagon Books).

Latour, B. & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts* (Beverly Hills, CA: Sage).

Lerner, D. & Lasswell, H.D. (1951). *The Policy Sciences* (Stanford University Press).

Leydesdorff, L. (1989). The relations between qualitative theory and scientometric methods in science and technology issues, *Scientometrics 15 (5/6)* 333-347.

Lievrouw, L.A. (1988). Bibliometrics and invisible colleges: at the intersection of communication research and information science, *Proceedings of the American Society for Information Science 25* 54-58.

Lievrouw, L.A. (1990). Reconciling structure and process in the study of scholarly communication. In: Borgman, C.L. (ed.) *Scholarly Communication and Bibliometrics* (London: Sage, 59-68).

Lievrouw, L.A. and others (1987). Triangulation as a research strategy for identifying invisible colleges among biomedical sciences, *Social Networks 9(3)* 217-238.

Lindblom, C.E. (1959). The science of muddling through, *Public Administration Review 19(1)* 78-88.

Lindblom, C.E. (1979). Still muddling, not yet through, *Public Administrative Review 39(6)* 517-525.

Ling, R.F. & Killough, G.S. (1976). Probability tables for cluster analysis based on a theory of random graphs, *Journal of the American Statistical Association 71(3)* 293-300.

Lotka, A.J. (1926). The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences 16(12)* 317-323.

McGarry, K. (1991). Epilogue: differing views of knowledge. In: Meadows, A.J. (ed.) Knowledge and Communication: Essays on the Information Chain (London: Library Association, 132-152).

Mahon, B. (1997). European information policy: the role of institutional factors. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series).

Malley, I. (1988). *National Information Policy in the United Kingdom* (Shepshed: IMPC).

Marshakova, I.V. (1973). Bibliographic coupling system based on cited references, *Nauchno Teknicheskaya Informatsiya Seriya 2(6)* 3-8.

Marshakova-Shaikevich, I.V. (1993). Bibliometrics as a research technique in epistemology and philosophy of science, *International Forum on Information and Documentation 18(3/4)* 3-9.

Marton, J. (1985). Obsolescence or immediacy? Evidence supporting Price's hypothesis, *Scientometrics 7(1)* 145-153.

Martyn, J. (1992). UK government policy on the electronic information industry, *Journal of Information Science 18(4)* 269-272.

Martyn, J., Vickers, P. & Feeney, M. (1990). *UK Information 2000* (London: Bowker-Saur).

McCain, K.W. (1990). Mapping authors in intellectual space: a technical overview, *Journal of the American Society for Information Science 41(6)* 433-443.

McIntosh, T.J. (1990). *Federal Information in the Electronic Age: Policy Issues for the 1990s* (Washington: Bureau of National Affairs).

McLain, J.P. (1990). Bibliometrics Toolbox, *Journal of the American Society for Information Science 41(1)* 70-71.

Merton, R.K. (1973). *The Sociology of Science* (University of Chicago Press).

Milevski, S. (1986). Information policy through public laws of the 95th-98th Congresses, *Proceedings of the American Society for Information Science Annual Meeting 23* 211-219.

Milligan, G.W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms, *Psychometrika 45(4)* 325-342.

Moed, H. (1989). Bibliometric measurement of research performance and Price's theory of differences among the sciences, *Scientometrics 15(3)* 437-483.

Moore, N. (1992). *Information Policy: a Discussion Paper on the Scope of the Policy Studies Institute* (London: Policy Studies Institute).

Moore, N. (1993). Information policy and strategic development: a framework for the analysis of policy objectives, *Aslib Proceedings 45(11/12)* 281-285.

Moore, N. (1997). Neo-liberal or dirigiste? Policies for an Information Society. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 89-100).

Moore, N. & Steele, J. (1991). *Information-Intensive Britain: A Critical Analysis of the Policy Issues* (London: Policy Studies Institute).

Mullins, N.C. (1968). The distribution of social and cultural properties in informal communication networks among biological scientists, *American Sociological Review 33(10)* 786-797.

Nadel, E. (1980). Formal communication, journal concentration and the rise of a discipline in physics, *Sociology 14(3)* 401-416.

Nagpaul, P.S. & Sharma, L. (1995). Science in the eighties: a typology of countries based on inter-field priorities, *Scientometrics 34(2)* 263-283.

Nicholls, P.T. (1988). Price's square root law: empirical validity and relation to Lotka's law, *Information Processing & Management 24(4)* 469-477.

Nora, S. & Minc, A. (1980). *The Computerisation of France: A Report to the President of France* (Massachussetts Institute of Technology Press) [English translation].

Norušis, M.J. (1994). *SPSS Professional Statistics 6.1* (Chicago: SPSS Inc.).

Oliu-Vukovic, V. (1991). The shape of the distribution curve: an indication of changes in the journal productivity distribution pattern, *Journal of Information Science 17(5)* 281-290.

Oppenheim, C. (1996). An agenda for action to achieve the information society in the UK, *Journal of Information Science 22(6)* 407-421.

Osareh, F. (1996a). Bibliometric, citation analysis and co-citation analysis: a review of literature I, *Libri 46(3)* 149-158.

Osareh, F. (1996b). Bibliometric, citation analysis and co-citation analysis: a review of literature II, *Libri 46(4)* 217-225.

Overman, E.S. & Cahill, A.G. (1990). Information policy: a study of values in the policy process, *Policy Studies Review 9(4)* 803-818.

Paisley, W. (1984). Communication in the communication sciences, In: Dervin, B. & Voigt, M. (eds), *Progress in the Communication Sciences, Vol. 5* (Norwood, NJ: Ablex, 1-43).

Paisley, W. (1990). The future of bibliometrics. In: Borgman, C.L. (ed.) *Scholarly Communication and Bibliometrics* (London: Sage, 281-299).

Parsons, W. (1995). *Public Policy: An Introduction to the Theory and Practice of Policy Analysis* (Aldershot: Edward Elgar).

Peritz, B.C. (1980). The methods of library science research: some results from a bibliometric survey, *Library Research 2(3)* 251-268.

Peritz, B.C. (1988). The literature of demography: its characteristics and changes over time, *Journal of Information Science 14(2)* 99-107.

Pierce, S.J. (1990). Discipinary work and interdisciplinary areas: sociology and bibliometrics. In: Borgman, C.L. (ed.) *Scholarly Communication and Bibliometrics* (London: Sage, 46-58).

Popper, K.R. (1959). *The Logic of Scientific Discovery* (London: Hutchinson).

Porat, M. (1977). *The Information Economy: Definition and Measurement* (Washington: Office of Telecommunications).

Price, D.J. DeSolla (1965). The science of scientists, *Medical Opinion and Review 1(10)* 88-97.

Price, D.J. DeSolla (1969). Measuring the size of science, *Proceedings of the Israel Academy of Sciences and Humanities 4(6)* 98-111.

Price, D. J. DeSolla (1970). Citation measures of hard science, soft science, technology and nonscience, In: Nelson, C.E. & Pollack, D.K. (eds), *Communication Among Scientists and Engineers* (Lexington, 1970, 3-22).

Price, D.J. DeSolla & Gursey, S. (1975). Some statistical results for the numbers of authors in the states of the United States and the nations of the world, In: *ISI's Who is Publishing in Science* (Philadelphia: ISI Press).

Pritchard, A. (1968). *Computers, Statistical Bibliography and Abstracting Services* (unpublished MS quoted in Borgman, 1990:13).

Pye, D. (1997). Changing the corporate culture: information policies for business excellence. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 191-205).

Quade, E.S. (1976). *Analysis for Public Decisions* (New York: Elsevier, 3rd edition, 1983).

van Raan, A.F.J. (1991). Fractal geometry of information space as represented by co-citation clustering, *Scientometrics 20(3)* 439-449.

Rahmann, S. (1996). Information policies for developing nations: a framework for analysis applied to Malaysian and Indian information policies, *Libri 46(4)* 184-195.

Raptis, P. (1992). Authorship characteristics in five international library science journals, *Libri 42(1)* 35-52.

Rees-Potter, L.K. (1980). Identification of areas of social science research in Canada: a bibliometric analysis, *Proceedings of the American Society for Information Science 17* 282-284.

Rennie, J.S. (1977). The social dimension of information transfer, In: De Mey, M. and others (eds), *CC77 International Workshop on the Cognitive Viewpoint* (University of Ghent, 219-224).

Rice, R.E. and others (1989). Journal to journal citation data: issues of validity and reliability, *Scientometrics 15(3/4)* 257-282.

Robinson, J.J. (ed.) (1982). *Policy Styles in Western Europe* (London: Allen & Unwin).

Rosenberg, V. (1982). National information policies, *Annual Review of Information Science and Technology 17* 3-32.

Rowlands, I. (1983). *Clustering Tendency in Document Test Collections: Cluster Validation In Automatic Document Classification* (University of Sheffield, MSc dissertation).

Rowlands, I. (1995). Towards public-private synergy in the European electronic information services market, *Journal of Government Information 22(3)* 227-235.

Rowlands, I. (1996). Understanding information policy: concepts, frameworks and research tools, *Journal of Information Science 22(1)* 13-25.

Rowlands, I. (1997). General overview. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 3-16).

Rowlands, I. & Turner, P. (1997). Models and frameworks for information policy research. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 46-60).

Rowlands, I. & Vogel, S. (1991). *Information Policies: A Sourcebook* (London: Taylor Graham).

Ruffin, R.J. & Gregory, P.R. (1990). *Principles of Economics* (Glenview, IL: Scott, Foresman, Little, Brown, 4th edition).

SATCOM (1969). *United States National Academy of Sciences SATCOM Committee, Scientific and Technical Communication: a Pressing National Problem and Recommendations for its Solution* (Washington: National Academy of Sciences).

Sheppard, E.D. & Bawden, D. (1997). More news, less knowledge? An information content analysis of television and newspaper coverage of the Gulf War, *International Journal of Information Management 17(3)* 211-227.

Shinn, T. (1982). Scientific disciplines and organizational specificity. In: Elias, N., Martins, H. & Whitely, R.D. (eds), *Scientific Establishments and Hierarchies* (Dordrecht: Ridel, 165-195).

Sillince, J.A.A. (1994). Coherence of issues and coordination of instruments in European information policy, *Journal of Information Science 20(4)* 219-236.

Small, H.G. (1973). Cocitation in the scientific literature: a new measure of the relationship between two documents, *Journal of the American Society for Information Science 24(3)* 265-269.

Small, H.G. (1977). A cocitation model of a scientific specialty: a longitudinal study of collagen research, *Social Studies of Science 7(2)* 139-166.

Small, H.G. (1978). Cited documents as concept symbols, *Social Studies of Science 8(3)* 327-340.

Small, H.G. & Crane, D. (1979). Specialties and disciplines in science and social science: an examination of their structures using citation indexes, *Scientometrics 1(5/6)* 445-461.

Small, H.G. & Garfield, E. (1986). The geography of science: disciplinary and national mappings, *Journal of Information Science 11(4)* 147-159.

Small & Greenlee, E. (1990). A cocitation study of AIDS research, In: Borgman, C.L. (ed.), *Scholarly Communication and Bibliometrics* (Newbury Park, CA: Sage, 166-193).

Sneath & Sokal (1973). *Numerical Taxonomy* (San Francisco: W.H. Freeman & Co.).

Spiegel-Rosing, I.S. (1972). Journal authors as an indicator of scientific manpower: a methodological study using data for the two Germanies and Europe, *Science Studies 2(4)* 337-359.

Spiegel-Rösing, I.S. (1977). Science studies: bibliometric and content analysis, *Social Studies of Science 7(1)* 97-113.

Stevens, R.E. (1953). Characteristics of scientific literatures, *American Council of Learned Societies Monograph 6* (quoted in Becher, 1989:87).

Stewart, R.K. (1990). *Access and Efficiency in the Reagan-era Information Policy: a Case Study of the Attempt to Privatize the National Technical Information Service* (University of Washington, PhD thesis).

Stokes, T.D. & Hartley, J.A. (1989). Coauthorship, social structure and influence within specialties, *Social Studies of Science 19(2)* 101-125.

Strachan, J.S. & Rowlands, I. (1997). Information for policy-making. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 61-73).

Stroetmann, K.A. (1993). Information research in Europe: an investigation, *Journal of Information Science 19(2)* 149-154.

Sullivan, D.; White, D.H. & Barboni, E.J. (1977). Cocitation analyses of science: an evaluation, *Social Studies of Science 7(2)* 223-240.

Suprenant, T.T. (1987). Problems and trends in international information and communication policies, *Information Processing & Management 23(1)* 47-64.

Tijssen, R.J.W. (1993). A scientometric cognitive study of neural network research: expert mental maps versus bibliometric maps, *Scientometrics 28(1)* 111-136.

Tijssen, R.J.W. and others (1990). Integrating multiple sources of information in literature-based maps of science, *Journal of Information Science 16(4)* 217-227.

Todorov, R. & Glänzel, W. (1988). Journal citation measures: a concise review, *Journal of Information Science 14(1)* 47-56.

Toulmin, S. (1972). *Human Understanding* (Oxford: Clarendon Press).

Trauth, E.M. (1986). An integrative approach to information policy research, *Telecommunications Policy 10(1)* 41-50.

Turner, P. (1997). Information policy concepts: overview. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 19-26).

UNESCO (1981). *Guidelines on National Information Policies: Scope, Formulation and Implementation* (Paris: UNESCO).

US Domestic Council (1976). *National Information Policy - the "Rockefeller Report"* (Washington: US Government Printing Office for the Committee on the Right to Privacy).

Vickery, B.C. & Vickery, A. (1987). *Information Science in Theory and Practice* (London: Bowker-Saur).

Webb, E.J. and others (1966). *Unobtrusive Measures: Non-reactive Research in the Social Sciences* (Chicago: Rand McNally).

Weinberg, A. (1963). *Science, Government and Information: The Responsibilities of the Technical Community and Government in the Transfer of Information* (Washington: Government Printing Office).

Weingarten, F.W. (1989). Federal information policy development: the Congressional perspective, In: McClure, C., Hernon, P. & Relyea, H. (eds), *United States Government Information Policies: Views and Perspectives* (Norwood: NJ: Ablex).

White, H.D. (1981). Cocited author retrieval online: an experiment with the social indicators literature, *Journal of the American Society for Information Science 32(1)* 163-171.

White, H.D. & Griffith, B.C. (1981). Author cocitation: a literature measure of intellectual structure, *Journal of the American Society for Information Science 32(3)* 163-172.

White, H.D. & Griffith, B.C. (1982). Authors as markers of intellectual space: co-citation in studies of science, technology and society, *Journal of Documentation 38(4)* 255-272.

White, H.D. & McCain, K.W. (1989). Bibliometrics, *Annual Review of Information Science and Technology 24* 119-186.

Whitley, R. (1976). Umbrella and polytheistic scientific disciplines and their elites, *Social Studies of Science 6(6)* 471-497.

Whitley, R. (1984). The establishment and structure of the sciences: reputational organizations. In: Elias, N., Martins, H. & Whitely, R.D. (eds), *Scientific Establishments and Hierarchies* (Dordrecht: Ridel, 141-157).

Winstanley, M. (1976). Assimilation into the literature of a critical advance in molecular biology, *Social Studies of Science 6(6)* 545-549.

Woolgar, S. (1991). Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy, *Science and Public Policy 18(5)* 319-326.

Worlock, D. (1997). Real policy or 'virtual' policy? A case study of tradeable information policy. In: Rowlands, I. (ed.) *Understanding Information Policy* (London: Bowker-Saur British Library Research series, 146-158).

Wouters, P. & Leydesdorff, L. (1994). Has Price's dream come true: is scientometrics a hard science? *Scientometrics 31(2)* 193-222.

Zinman, J. (1994). *Prometheus Unbound: Science in a Dynamic State* (Cambridge University Press).