AN ASSESSMENT OF STEREOTYPICAL MODELS OF ON-LINE SEARCHING BEHAVIOUR: END-USERS

CASE STUDY: PRACTITIONERS
(POLITICIANS AND JOURNALISTS)

David A. R. Nicholas

**Submitted in fulfilment of the requirements for the degree
of PhD (Information Science)**

**Department of Information Science, School of Informatics
CITY UNIVERSITY**

September 1995

CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

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.

David Nicholas

ABSTRACT

The purpose of this study was to determine what the searching characteristics of end-users in a non-academic environment (practitioners) were and to discover whether these characteristics were those that are ascribed to end-users in the professional literature and folklore, and whether they differed materially from those of information professionals working in the same fields. Searching characteristics were interpreted in their widest sense to include: volume and frequency of use, when searches were done, file selection, use of system commands, types of search, searching strategies, searching behaviour, user satisfaction, obstacles to searching, willingness to delegate and levels of training. These characteristics were related to the information needs and information seeking behaviour of the end-users involved. These issues were explored in relation to two practitioner groups - journalists at *The Guardian* newspaper, and politicians at The House of Commons. Comparisons were made throughout with the respective Library staff in these two institutions.

A mixture of social and statistical methods were used to monitor end-user and professional searching in the two environments examined, though much of the analysis relies upon transactional log analysis. The period of study largely covered the years 1989-1995, though historical data for the journalists went back to the mid-eighties.

The principal findings of the study were as follows: politicians and journalists had much in common in respect to their on-line searching behaviour; in some respects end-users did conform to the picture that information professionals have of them: they did search with a very limited range of commands, more of their searches produced zero hits, and search statements were simple in construction. But in other respects they confound that image - they can be very quick and economical searchers, and they certainly did not display metres and metres of print-out. However, there were big variations between individual end-user groups, and it was often possible to find a end-user group that matched a information professional group on one aspect of on-line searching or another. The on-line behaviour of end-users was very much related to two factors: (1) their general information seeking behaviour; and, (2) the fact that they were not trained. There was a strong desire on the part of end-users to do their own searching and large volumes of searching were achieved, despite the fact that both groups of end-users had big obstacles put in their way: the twin high hurdles of poor access and limited training. Finally, great differences were found in the way in-house and dial-up users searched on-line systems.

GLOSSARY

In describing on-line systems and their users a number of often repeated terms have been used which require definition and clarification.

End-user	*A user of a library or information service who searches on-line databases themselves.*
Intermediary	*Someone who searches databases on behalf of others, typically, but not always, a professional librarian.*
Librarian	*A professionally qualified librarian or someone employed exclusively to do library work.*
Library researcher	*Subject specialists working in The House of Commons Library, who provide MPs and their Assistants with oral and written advice.*
Member	*A Member of Parliament.*
Practitioner	*A member of a profession who is employed to practice rather than teach the subject in question i.e. a Solicitor rather than a law lecturer.*
POLIS	*Parliamentary on-line information system.*
Research Assistant	*A member of an MPs personal staff who is largely responsible for finding and packaging information.*
User	*Someone who uses information rather than someone who has a responsibility for its storage and dissemination. All end-users are users, but not all users are end-users.*

CHAPTER 1
INTRODUCTION

1.1 Rationale

The original idea for the study arose out of the author's investigations of the impact of on-line systems on strategic groups of social science practitioners (Nicholas *et al* 1987; 1988). While conducting these studies librarians would frequently come to the researcher with stories that disparaged the on-line searching skills of their end-users. No hard evidence would be provided in support, but many anecdotes would be recounted. There was always the suspicion in the mind of the researcher that the librarians were saying this because they were unhappy with the whole idea of end-use; for they felt threatened for their (alleged) superior searching skills were now being put to the test. The driving force then behind this dissertation was to determine whether such blanket descriptions of end-user searching were in fact accurate.

An examination of the published literature shows that the librarians' views and concerns were mirrored in the literature, though a good deal of it is contradictory in nature. So many things have been said in the literature about the use (and non-use) of on-line services by end-users that it is easy to become confused, and uncertain as to what is fact, fantasy or hearsay. And this is very much more the case with those end-users working outside the academic world (practitioners) - the focus of this study, about whom there had been much rumour but only limited research. However, a key theme running through much of the discussion is that, indeed, end-users are infrequent, poor or limited searchers. Oppenheim (1992: 27) summarised it so: *end-users are not necessarily aware of on-line, or even if they are, they need convincing that they actually should use it; they do not understand the plethora of on-line hosts, of the databases and of the commands needed to log on, do the search and log out. Even if they did use on-line, they would not use it much - perhaps once a week ... .* And as far as Barker (1993) is concerned the moment you let users do their own searching quality goes out the window. Larsen (1988: 468) extends the line of critical fire: *we can assume that ... 35-50% of the searches* [of end-users] *will result in nothing, and that even when material is found it is not always relevant.*

Very much part of the Professional folklore, too, is that, while end-users are quick and dirty searchers, information professionals are certainly not. Indeed, information professionals owe much of their on-line existence to the (largely) self-proclaimed belief that they are highly skilled and effective searchers. The argument goes: that because of their professional training, the greater opportunities that they have to practise their on-line skills, and their professional commitment to improve and update their skills; their searches are superior to those of end-users. The professional literature - no doubt with a little self-interest and self-preservation in mind - is rich in studies parading the professional searchers' virtues (Lancaster *et al.*, 1994). If all this is, indeed, true why then are users taking to database searching in ever increasing numbers, and more pertinently, perhaps, why when asked, do end-users seem so satisfied with the products of their own searching (Jacobson and Ullman, 1989)?

These beliefs about end-users are very much built into information system design and are also at the heart of the end-user training and marketing strategies of the on-line hosts. Thus FT PROFILE, despite having what arguably is the most end user-friendly software of all on-line hosts (Harold Jackson *in* Nicholas, D *et al.* 1987: 59), have launched a succession of (allegedly) even more friendly interfaces (Freeway, INFOplus and, most recently, FT Discovery). DIALOG too, in the hunt for the elusive and financially attractive end-user, continues to bring out menu driven variants of its successful command-driven files. The trouble is that it is very difficult to reach the truth about end-use because, seemingly, everybody with a modicum of interest wants to get in on the act, and in so doing, frequently muddying the waters with statements based on small scale, limited duration studies. There certainly does appear to be an open season on end-users in the professional press. Larsen (1988), in fact, believes that many of these statements are at best educated guesses, based on very little evaluative work. Borgman (1986: 387) echoes Larsen's concerns about existing end-user research, suggesting that *we do not have yet sufficient knowledge of user behaviour to make major improvements in systems design and training.*

There are good grounds for believing that more research needs to be conducted into the on-line endeavours of practitioner end-users. The grounds for thinking so are these:
- Practitioners have very different information needs to those of academics and seek information in different ways. This was shown to be very much the case by the Investigation of the Information Needs of Social Scientists and the Design of Information Systems in the Social Sciences research projects of the late Sixties and early Seventies. (Bath University, 1971). End-use needs to be seen and explained

in this broader context of information need. The special interest in studying practitioners as end-users lies in the fact that: (a) they are not generally supplicants in the information seeking process or beholden in anyway to the information professionals - as many academics are, for this can lead to users conforming to the expectations of the information professionals, who inevitably conduct the studies; (b) information seeking is not a separate activity and is often a normal and routine aspect of a busy job - and conducted at the office desk, as it is, for instance, in the case of journalists.

- In the absence of more substantial research, the on-line characteristics of academics, about which we know an awful lot, will also be ascribed to practitioners. Indeed, much of what we know about end-users comes either from studies of academics - a tame, but highly unrepresentative group: of practitioners, a potentially much bigger and arguably more important group, we know relatively little. The present study hopes to restore some of the balance.

- Until quite recently end-users simply did not have sufficient exposure to on-line systems to be able to establish patterns of on-line searching. Therefore the results of many of the earlier studies questioning the abilities of end-users are plainly in need of review. However, a number of end-user groups - especially journalists, have now had unfettered access to on-line systems for some years now, and in many cases searching appears to have become very much part of their daily work routines.

- The problem is compounded by the fact that end-users are far from being a homogeneous group (far more diverse than information professionals for instance - their size and relative anonymity dictates that anyway) and generalisations are always going to be questionable: what is possibly true of one group is unlikely to be true of others.

- So-called comparative studies with intermediaries have not always been conducted on a level playing field, typically end-users have had poorer access to on-line systems than the intermediaries. Rubbishing the abilities of end-users on the basis of restricted (and formative) searching is plainly wrong.

- The fact that things in the information world are rapidly changing also adds an element of uncertainty to the whole end-user debate. Users are becoming more familiar with computer paraphernalia - through massive exposure if nothing else

(to systems like the Internet), and the need to develop information searching skills amongst the various professional groups - such as nurses - is increasingly being recognised. Not only are the demarcation lines between end-users and information professionals becoming increasingly blurred but the same too is happening with the once different types of on-line system: fancy front ends mean that the same information system can be offered to a whole range of users with various levels of skill or experience: this is especially the case with CD-ROM - the new information system *wunderkind*.

The researcher has also another personal reason for undertaking the research. Having taught on-line searching to library school students for many years and observed them coming to terms with on-line searching - frequently for the first time - it has always struck the researcher how difficult most of the library school students find it - even to move from searching to display mode and back again, something which typically involves no more than three commands. Yet search sessions are preceded by lengthy explanations of the commands and detailed handouts are provided well ahead of searches: and for some of these students (the part-timers) this is not the first time they have used an on-line system (although probably not the particular system being studied). It seems that, to begin with anyway, no on-line system is that user-friendly, nor do (novice) information workers bring a lot of knowledge with them. Also, students develop their skills at a very different pace. If this is the case with trainee information professionals why cannot it be true of end-users too? So why then consign all end-user searching to the dustbin?

1.2 Aims and objectives

The prime task of the research is to examine the use of on-line database systems by practitioner end-users and to explain this in the light of their information needs and general information seeking behaviour. Hopefully, providing a more considered and sympathetic picture as a result. Then, on the basis of this data, determine whether their use of on-line conforms to the stereotypical picture found in the published literature and professional folklore, and whether it differs from what has been discovered about academic end-users. Inevitably, this also involves assessing the conventional and principal method of dividing-up on-line users into intermediaries (professional information workers) and end-users and seeing whether it makes any sense to classify on-line users in such a way. Are these, once useful, labels for categorising on-line user types valid anymore; have time and technological change rendered them obsolete? Does such labelling point to significant, meaningful and

consistent distinctions in searching behaviour that can be put to good effect in information system design? Or, as has been suggested earlier, is it just a convenient, but crude, way of clumping together, what are essentially large and disparate groups of on-line users?

The dividing principle behind the information intermediary/end-user classification is degree or level of (information) professionalism, or perhaps more accurately in today's increasingly deregulated information environment, whether on-line searching is the person's principal job or not. Behind it all though lies the often unstated, but implied assumption, that, as a consequence, professionally trained searchers are more expert and more demanding (of systems). It is perfectly understandable that such a distinction between on-line users should be made, given: (a) the higher levels of training and education information professionals obtain in on-line searching and information seeking in general; (b) their greater opportunities for going on-line - after all the typical intermediary services the on-line needs of not just one, but many; and (c) the greater motivation to learn and enhance skills that might be expected from individuals actually working in the information profession. However, that does not necessarily make it the only, or best way, of categorising users - after all both groups are far from homogeneous; one might expect that even information professionals would vary in their on-line expertise, levels of professionalism, motivation and experience (Richardson, 1981).

However, even if all this follows, does it necessarily mean that intermediaries provide the better search results: does a detailed knowledge of system commands and long hours on-line necessarily mean that the product of the search is actually superior? Surely, what makes a good search, is crucial to the whole debate - and at the end of the day the only opinion really worth hearing is that of the consumer - and when asked about their searches they say that they are generally well satisfied with what they get from their on-line labours (Jacobson and Ullman, 1989).

While the specific aim of the study is to investigate the nature of end-use amongst practitioners it is not possible to do this without investigating the way that intermediaries in the field search, too. For, after all, the criticisms of end-users are usually made with intermediary searching behaviour in mind. Thus, for instance, to know that an end-user takes, say, four minutes to complete a search means little on its own: the real interest probably lies in whether this is longer, quicker or the same time as that taken by an information professional.

1.3 Scope/definition of study area

1.3.1 End-user groups featured

Given the potentially huge field of investigation, the aims and objectives of the study can only be sensibly met by concentrating on a small number of professional groups. To provide the research with the necessary weight, depth and width, the subject fields to which the practitioners belonged had to boast substantially sized on-line using groups, who were provided with (potentially) good on-line access, could be surveyed closely for a period of months or years, rather than days; and, what is more, had a group of information professionals who could provide the essential comparisons. A tall order, because such groups are not easily found and studied - hence the attraction of academics to information researchers, but two have been identified for this study: politicians (MPs and their political staff) at The House of Commons and journalists at *The Guardian*. Both groups have been researched by the author for some years now - ten years in the case of journalists and two in the case of politicians.

There are a number of very good reasons for choosing these two particular groups. Let us deal first with the case of national politicians and the reasons for selecting them.

- There are strong methodological grounds for selection - and this is simply not just a case of the tail wagging the dog. At the House of Commons the on-line searches of all types of on-line user have been minutely and routinely monitored by computer, offering up what must be one of the largest collections of on-line use data relating to practitioners: something substantial enough to compare with the academic OPAC studies.
- Through The House of Commons data the activities (and non-activity) of several hundred individuals could be investigated - a number about which you can begin to make meaningful statements.
- The House of Commons offered not just a sizeable body of end-users, but also, head-to-head comparisons with information professionals. Additionally, users at The Commons were highly defined and demarcated in educational, administrative and social terms, a stratification which allowed for comparisons of well defined sub-groups of online users. Essentially there were four distinctive groups of on-line user at The Commons: (1) Members of Parliament; (2) their Research Assistants (generally academic or political researchers, with little or no information training); (3) Subject specialists employed by the Library to offer verbal and written advice and research to Members; (4) the professional librarians working in the Library. The whole spectrum of on-line user groups were thus on show.

- At The Commons end-users (Members and their Assistants) did not have to conduct their own searching: they could always delegate the search to highly trained Library staff.
- They had access to two very different systems - a bibliographic one (the Parliamentary Online Information System, POLIS) and a full-text one (TEXTLINE). Not only could the alleged preference for full-text systems be tested, but it was also possible to see if searching styles differed, and whether end-users really could not cope with more than one host at a time.
- MPs and their assistants have had their principal on-line resource - POLIS - customised to meet their own special needs: they searched POLIS through menus, largely from dedicated terminals, though Library staff used the original command-driven version. Menus are supposedly the end-users' friend, so here was another interesting variable to investigate.
- There was the relevance of on-line data for MPs and their staff, and the urgent need for it. MPs are expected to be informed about a whole range of topics, and often at very little notice (the media may call up at any time for an interview, a quote or a reaction to a major event). They also need to keep track of business in the Commons, check what they have said in the past and discover what opposing MPs and Ministers have said.

Journalists introduce the essential comparative element to the study. And subject studied or practised is possibly an important variable in on-line searching behaviour. Though often closely connected with politicians and interested in all things political, journalists do present a somewhat different end-user case.
- They inhabit a far less orderly and centralised environment.
- Information is their lifeblood - some journalists are in fact nothing more than information processors and packagers.
- They are also far more buccaneering and independent in their information approaches and they are much more practised information seekers too (possibly, together with economists, the largest and busiest group of end-users around in the social sciences). Unlike most politicians they typically search from their desks. If anything, their demands for currency and speed of delivery are even more stringent.
- They are generally conversant with, and surrounded by, on-line technology.
- And, of course, journalists (and politicians for that matter) are anything but supplicants when it comes to their use of information systems. Both groups feature high-profile individuals of a strong, (sometimes) arrogant and independent mind.

Amongst journalists those from *The Guardian* commend themselves most strongly as a subject for end-user study.

- They have been probably using on-line as long, if not longer than any other practitioner group - through them it is possible to examine more than a decade of end user searching.
- They are generally high volume users.
- They have direct and largely unfettered access to a full-text system (FT PROFILE) of potentially immense worth to their work - it has their own newspaper on it for instance.
- Librarians' searching largely revolves around the same on-line system, enabling level-playing field comparisons to be made.
- High degrees of co-operation were forthcoming thanks to good relations with both management and library staff.

A note is required here to explain the nomenclature used to refer to user groups throughout the thesis. The word end-user itself requires explanation because it can mean both users and potential users of an on-line information service. The same is true of the term user when used in a library services context. This study concerns itself primarily with searching behaviour, therefore, in the great majority of cases it is active end-users that are being considered. Having said this though, some politicians and journalist who did not conduct their own searches were questioned to determine the reasons for non-use and delegation.

Practitioners is the preferred term to describe non-academic users - the group on which this study focuses. It refers to professional groups who practice rather than teach a particular subject. e.g. Solicitors, stockbrokers, journalists, politicians. This term first obtained prominence and professional acceptance with the work of the British Library Research and Development Department (BLRDD) funded Investigation of the information requirements of social scientists (Bath University, 1971: 14). It has since been taken up by other writers (Brittain, 1992: 2). Because of the nature of the two groups selected for study it is largely social science practitioners who are being referred to. The other distinction that is made throughout the study is that between end-users (a practitioner is an end-user) and information professionals or intermediaries. This is a distinction between someone who searches an on-line system as part of doing another job (an end-user) and someone whose principal role is to search on-line systems on behalf of others - either a professionally trained librarian or someone employed by the library. There is a complication, and this arises in the case

of The House of Commons. Firstly, the principal group of end-users studied there - MPs Research Assistants - also functioned as intermediaries (to MPs). Secondly, the Library staff at The Commons also included a group of users who were not professionally trained librarians - these were the Library's research staff. Nevertheless, although the special status of these two groups is taken into account in the analysis, for convenience and economy these groups are, respectively, referred to as end-users and librarians.

1.3.2 Types of on-line system investigated

Full-text systems have an obvious attraction to end-users, for they are a one-stop information service, providing information in a relatively familiar format - a major attraction, particularly for deadline-driven journalists and busy politicians. Politicians at The House of Commons had access to TEXTLINE - more accurately described as a near text system, and journalist had access to FT PROFILE. Both are remote dial-up commercial hosts. It is in the use of such systems by end-users that research is lacking. However, the study did not simply confine itself to full-text services, for politicians at The House of Commons, also had access to what might be loosely described as an OPAC, POLIS. In fact, because POLIS's computer monitoring was much more extensive and detailed than TEXTLINE's, POLIS features more strongly in the analysis of politicians' on-line searching.

The principal on-line systems that feature in the investigation of the two case study groups - POLIS and FT PROFILE - are in fact unlikely bedfellows, the former being an essentially locational tool, not unlike a library OPAC (although POLIS can be accessed commercially and remotely), and the latter being a commercial full-text, end-user service *par excellance*. From a research point of view these large differences between the systems provided an ideal test-bed against which to assess the various claims made against end-users. The fact that this particular selection also threw up: a dial-up service (TEXTLINE and PROFILE); and an in-house service (POLIS); a command-driven service (PROFILE) and a menu-driven service (POLIS); lent a further dimension to the study.

1.3.3 Characteristics of end-use studied

The chief stereotypical traits of end-users that required investigation were these: levels of searching skill and command knowledge - end-users were alleged to be deficient here (Arundale and Erbach, 1988: 577; Stanbridge, 1992); the amount and frequency

of searching - end-users were reputed to be infrequent and/or low-volume users (Walton, 1983: 49); familiarity with sources - many end-users were said to lack source knowledge (Oppenheim, 1992: 27); searching style/approach - thought to be simple and broad and characterised by much browsing (Rigglesford; 1992); the success and satisfaction with, searches - there were thought to be many failures (Peters, 1989).

More specifically the following on-line characteristics will be examined:

- Levels of activity. The volume of searching conducted: considered variously from the point of view of the number of sessions/searches conducted, the frequency of searching, the amount of time spent on-line, the number of records/screens displayed and the number of individual transactions conducted.
- Duration of searches.
- Time of searching - hour, day and month.
- File selection and utilisation.
- Command utilisation The range and type of searching commands used. Special attention given to the use of so-called *advanced* commands: truncation, Boolean commands, proximity operators.
- The types of search undertaken - subject (default/broad/narrow), author, date etc.; simple or complex (as shown by number of stages in a search and number of terms employed).
- Browsing and displaying characteristics.
- The productivity and success of searches, covering posting counts, items displayed (especially full-text ones), problems encountered (failure rates, errors made).
- Levels of user satisfaction.
- Attitudes towards on-line searching.
- Reasons for searching and delegating.
- Training.

No analysis of on-line information seeking behaviour is complete without a consideration of the information needs and general information seeking behaviour of the subjects studied. Without such an understanding it is very difficult to interpret the raw data, that transactional log analysis in particular, generates in vast quantities. End-use is not conducted in a vacuum. It is not simply just a case of discovering, say, that journalists take 4 minutes to conduct a search - for that is the easy bit, but also determining **why** that should be so. In recognition of this the on-line searching behaviour of journalists and politicians is viewed, wherever possible, in the light of what is known about their information needs and information seeking behaviour.

1.4 Duration of study and source note.

The period of registration for this research project encompasses the years 1990-1995. However, the author has been investigating practitioner end-users much longer than that. The research line, in fact, stretches back as far as 1985. Therefore, incorporated in the current study is material collected before 1990, but not previously published in its present form. For reasons of flow, convenience and digestion, previously collected source material on the exact line of the current research project has been treated as an outcome of the current research project. This work largely emanates from the BLRDD funded Information Seeking in an Information Society research project, whose findings were published in two books: *Online searching: its impact on information users* (Nicholas *et al*; 1987) and *End-users of online information systems: an analysis.* (Nicholas *et al*; 1988). Similarly, material emanating from the current study and published after 1990 has also been treated as an outcome of the present study - a complete list of these publications can be found after the list of References. Data gathering for this thesis stopped in May 1995.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

There is a vast end-user literature - much of it emanating from the early 1980s, when end-user studies of all persuasions took hold of a profession worried at the impact that this new phenomena was going to have on their jobs. A quote from a *Financial Times* business report (Holmes, 1985) on the on-line industry gives a flavour of much of the writing at the time: *The emphasis is expected to shift further away from the librarian and researcher to the end-user because of the current explosion of microcomputer installations and increased user friendly software.* Many years later and a lot of the end-user literature still concerned itself with the spread of end-use and its implications. Thus Meadows and Buckle (1991) claimed that the information seeking behaviour of scientists was currently undergoing changes, with increasingly more of them conducting on-line searches themselves. In 1986 28% of their respondents searched on-line databases, whereas in 1991 that figure had more than doubled to 58%. Pursuing a similar line of enquiry, Barbuto and Cellavos (1991) conducted a questionnaire study in New England and found that 85% of the organisations contacted had end-users. Continuing the information profession's preoccupation with the inexorable rise of the academic end-user, East and Tilson (1993) point to factors such as increased familiarity with information technology, greater availability of databases and increased student numbers as driving growth the growth of end-use even higher. They also noted that the only area were librarians were the principle on-line users was in the case of commercial on-line hosts. In a document packed with statistics charting the rise of end-use, the nature and quality of searching got a bare mention - just something to consider in the future (Ibid.: 46).

When it comes to the published literature on practitioners as end-users much of it is light on substance and anecdotal in tone: typically short-term studies on very small groups, sometimes conducted in the writer's own organisation (Poynder, 1993; Stanbridge, 1992 etc.). Furthermore, much of the literature on practitioners concerns scientists or engineers (Richardson, 1981; Walton, 1983; Poynder, 1993) - the early on-line leaders, but groups that have very specialist - almost academic - information needs. By contrast, the literature on academics, even in the early 1980s, was extensive and detailed (Bodtke-Roberts, 1983; Bonta, 1983; Fjallbrant *et al*, 1983)

With so much research conducted on academics and scientists there is a real possibility that the on-line experiences of academics and scientists will be ascribed to social science practitioners groups, just as the information needs of scientists and academics were ascribed to social science practitioners, until the INFROSS study conducted in the late 1960s/early 1970s, showed that their needs were quite different. (Bath University, 1971)

The published literature on the methods associated with end-user investigations is discussed more fully in Chapter 3, the methods section.

2.2 OPAC and CD ROM research in an educational environment

Even though the writers' principal concern is with practitioners it would be wrong to neglect the sizeable literature on academic end users, for this is where most of the substantial work on end-users and their searching habits and styles has been done. It is also the area to which we have to look for methods, as the academic environment has proved a rich test-bed for the development of suitable methodologies, in particular transactional log analysis, a methodology that features strongly in this dissertation. Furthermore, as one of the writer's hypotheses is that, because of differences in information need, the nature of end user searching is possibly different in the practitioner environment, the academic studies have an important comparative and benchmark role to play in this thesis.

OPACS, because they are directly accessed by end-users without professional help, have been the subject of much research, especially by staff of City University's Department of Information Science. Researchers at City have been particularly concerned with the methods of mapping and evaluating searching behaviour at the OPAC terminal. In one prominent piece of research, interviews, questionnaires, observation, talk-aloud and transaction log analysis were all employed to obtain a comprehensive picture of searching behaviour (Hancock-Beaulieu et al., 1990B). The fact that the sample group were students from the same institution, and the closeness (to the researchers) and homogeneity of this group raises doubts about generalising the findings, but then this applies to many of the academic OPAC studies. The authors argue that the transaction log facility was not a wholly effective method for obtaining a holistic view of user behaviour at the catalogue and it had to be supported by other methods. On a more general front it was discovered that: (1) end-users' information needs are more nebulous than indicated by search terms or expressed topics (hence the need to delve deeper into the information needs of end-users); that the searcher

changes topic depending on what they find in the catalogue, and; (3) that refining searches are often made on an *ad hoc* or intuitive basis.

Hancock-Beaulieu's (1990A) principal concern has been with the on-line library catalogue, hypothesising that the technology has made more users interested in the subject approach - something that the early OPAC designers had not anticipated. In her study, based on a sample of 204 student manual and OPAC searches, it was found that: (a) subject searching was universally popular with both manual and on-line catalogue searchers, but, in fact, even more popular with manual users - 72% as opposed to 60% (thus disproving the initial hypothesis); (b) just over one third of all specific (known) item searches turned into subject searches during the search process; (c) hybrid (mixed) searches were much more a feature of the on-line environment - 37% of searches as opposed to 22% in the manual environment; (d) initial subject search expressions were very economical - averaging two terms; (e) 81% of the terms chosen to interrogate the OPAC coincided with access points - another 15% partly matched, however when on-line only 29% proved satisfactory with 52% of students narrowing their terms; (f) when the student's initial expressed topic was compared with their exit search, over three quarters had changed terms, with more narrowing than broadening their earlier selection - this being explained by student attempts to place their needs in a broader context to improve their success with the system; (g) far more subject searching and sifting took place at the shelves afterwards. Her conclusion was really, that the OPACs that existed at the time were ineffective as a support for the subject searcher - something that has particular relevance for the researcher's study of politicians (see Chapter 4).

Akeroyd (1990) has conducted a lot of work on end-users and their use of OPACs and CD-ROMs. One such study, featuring three different types of OPAC, involved final year university students (though the author does his best to avoid the term student), who had some familiarity with searching the systems. A very small numbers of students were involved - just 10, in fact. The principal findings of this study were: (1) the (rather obvious) fact that users, when given a choice of approaches, opted for the simplest search key - the one they could grasp most easily; (2) users had a lot of perseverance when searching : they tried to ratify a nil response by using alternative approaches to the same document (practitioners are probably unlikely to have the time to do this); (3) the more experienced users of the search party accepted zero results more readily; (4) surprisingly perhaps, users seemed to prefer controlled language subject searching rather than taking subject keywords from the title, but the author then goes on to belittle this result by claiming that this might have something to do

with the artificiality of the exercises set; (5) limited use of search vocabulary - with users hanging on to terms given to them in the exercise (the dangers inherent in the methods used); (6) users were satisfied with a small number of pertinent documents - perhaps using these as keys to unlock more material at the shelves. These results were buttressed by computer logs from the various OPACS involved - this produced more interesting results. It was found that most of the searches were known item ones, though this cannot have been surprising in a place where many users were given set readings. This is what really sets academic studies apart, users at the OPAC are typically chasing books and not information - as is the case with practitioners searching databases from their desks. Akeroyd also pointed to the large amount of browsing that went on, but didn't really quantify it, though he identified two types: compensatory browsing where because of some error or problem at the input stage the user tries to rectify the position by scanning large numbers of documents; and meaningful browsing - structured, directed and purposive searching. While an interesting distinction it must be very difficult to make the distinction, given the hurly burly of most searching. It was a distinction that could not be made in the logs supplied to the researcher.

Peters (1989), in large and impressive study of American academics, analysed the logs representing some 36,000 transactions over a period of 202 days. The major findings were that: (1) many searches failed - as many as 40% produced zero hits; (2) the major problems users encountered were to do with searching inappropriate databases (attributed to 39% of the problems), typographical and spelling errors, which together accounted for 21% of the problems; (3) title keyword searching (44% of all searches) was preferred to subject heading searches (16%); (4) low use of advanced features - 97.2% of searches were simple author, title or subject searches, with virtually no truncation being employed at all; (5) users exhibited a tendency to put all the information they had into a search, rather than hold some back (this concept of holding terms in reserve is a novel one); (6) nearly three quarters of users were either somewhat or very satisfied with their most recent search session; (7) many users tended to work with poor, inefficient high-recall searches; (8) a reluctance on the part of users to read and apply the help and guidance provided on the screens.

Barbuto and Cellavos (1991) in a similar type of investigation to that conducted by Peters, but using instead questionnaires as the prime research instrument, uncovered some additional and interesting facts about the searching techniques of end users. Firstly - and this too largely sets academics apart from many groups of practitioners, they all said that they wanted to conduct comprehensive searches, but what they

actually meant by a comprehensive search differed considerably - a third believing that 1-10 items retrieved would constitute such a search, while a quarter though this to be more than 50 items (more a librarians' definition of comprehensive, one would have thought). Secondly, they found that users were satisfied with their searches. They measured satisfaction by comparing the number of documents the users said they wanted, with what they actually found. To probably nobody's surprise those only requiring 1-10 documents were most satisfied. Thirdly, thesaurus use did not lead to higher levels of user satisfaction, non-users of the thesaurus retrieved more relevant documents; nor could users distinguish between descriptors and keywords (natural language terms).

The main trouble with transaction log studies is that they tend to contradict each other - it is not so much that different things are read into the data, more worryingly they sometimes come up with quite startlingly different results. There does however appear to be a general consensus about one thing, and that is that end-users get by with very few commands and ignore or evade such advanced facilities as Boolean searching and truncation. (Tolle, 1983; Peters, 1989). Thus Newkirk and Jacobson (1993) discovered that a quarter of all end-users searches contained no Boolean operators. Where Boolean operators were employed it tended to be the **and** that was used. The **or** was generally wholly neglected. Kirby and Miller (1986) found that 58% of search statements employed the **and**, but only 2% the **or**. Tenopir *et al* (1990) in an examination of the use of full text files also noted that **or** was used infrequently. Look much beyond this and the consensus begins to crack. For instance, in a surely 'rogue' study of ERIC searching, Bourne *et al* (1974) found that the **or** operator was used the most - twice as much as the **and** operator.

The use or non-use of truncation has fascinated researchers. Truncation is, of course, a convenient and camouflaged form of the Boolean **or,** though a concept few end-users appear familiar with. Akeroyd (1990) and Peters (1989) both found low or non-existent levels of truncation in their studies. Illustrating the contradictions inherent in log studies or, perhaps, the differences between on-line systems, Tenopir *et al* (1990) however, produced figures to show that truncation was used quite frequently by full text users - the full text and natural language nature of the database explaining the discordance?

A number of researchers (e.g. Ensor, 1992) have been troubled by another possible complexity inherent in searching OPACS, the use of controlled languages. In an extensive (and exhausting) questionnaire study, unfortunately full of indigestible

statistics, Ensor (predictably) discovered that, indeed, keyword knowledge was woeful, and that a knowledge of keywords (and Boolean operators) made users more satisfied with their searches. Barbuto and Cellavos (1991) as we heard threw a spanner in the works by declaring that thesaurus use - something quite important to controlled language searching - did not lead to higher levels of user satisfaction, non-users retrieved more documents; and that users commonly failed to distinguish between descriptors and keywords (natural language terms). Larsen's (1988) work points to end users moving the way of natural language searching.

Most transactional log studies have treated end-users as one big homogeneous group, which you can probably get away with rather more in the case of academics than you can with practitioners - a more heterogeneous group. However, when they have been broken down into groups according to age, gender, subject studied, mode of access or level of expertise interesting differences in searching behaviour have emerged - and this very fact has influenced the design and direction of this study. Undoubtedly, level of expertise has interested researchers most - after all this lies behind the whole end-user v intermediary debate. Gathering together the findings of two studies (Fenichel, 1979; Penniman, 1982) it seems that: frequent users employed more search commands (twice as many) as infrequent users and that they were quicker in conducting their searches. However, they are not noticeably more successful at searching nor did they make less errors.

Lancaster *et al.* (1994), in a highly relevant paper, looked specifically at the issue of end-user verses information professional searching. The main aims of the research project were to establish whether the searching of academic librarians produced better results than their users, and to compare their search approaches. The hypothesis being tested was, that experienced and trained searchers (the librarians) would get better results than novices (end-users) - despite the fact the end-users were best placed to determine what they wanted. Thirty five test searches were undertaken and, indeed, the librarians proved more successful in retrieving items that the end-users thought were useful. But the authors did point out that this has to be qualified by the fact that, even in the cases of the best professional searches, only half the items retrieved proved useful - in the end-users' case this was nearer one-third. This was attributed to the limited information the librarians had to work with (i.e. the request forms). A weak argument this, for many librarians, like those offering essentially phone-based information services, have less than that to work on. The relatively poor performance of the end-users was judged to have come about, not from the poor/limited use of search commands - like Boolean operators, but from an inadequate and limited search

vocabulary: in the words of the author: *they search too literally* (Ibid.: 375). The researchers also looked at the relative time taken to conduct the searches. End-users took on average twice as long to search: librarians averaging 20-25 minutes and end-users 55 minutes (quite long-times when compared to dial-up searching). Interestingly, there was no obvious correlation between length of search and performance in respect to success in obtaining relevant documents.

The authors also bemoan the subjectivity of much of the material published on end-user searching, claiming that it was largely based on user impressions, rather than objective evaluative data - however they do not make any reference to the rich body of OPAC research done in the UK, and already discussed here. What perturbed them most was the high levels of satisfaction expressed by users, when searches were plainly poor in recall and precision terms. This might well be the case, but on the basis of their own evidence, which they conveniently seem to ignore, in terms of the precision of their searches the end-users out-performed the professionals, and, in the case of recall, the librarians themselves only recorded a 46% success rate (Are they then to be chastised for their satisfaction with their searches?). Anyway, as far as the user is concerned, precision probably rates more highly than recall - a small number of highly relevant items is probably the best reward for most people, other than academics and researchers.

In regard to differences in searching behaviour resulting from gender (something which is briefly touched upon in the study of journalists), women have been found to use more commands than men and - possibly related - they spent longer on a session (Barber and Riccalton, 1988; Borgman, 1991). Of interest to this study too, differences have also been found to exist between in-house and dial-up users of the same system. (Sloan, 1986). Peters *et al* (1993) believe this to be a major influence on searching behaviour.

Treading a careful path between some of the contradictions in these OPAC studies, what they appear to tell us about academic end-users is that:
- they largely search on subject
- they have a small and limited search vocabulary
- a high proportion of their searches seemingly end-up with nothing in the way of documents (or success, if some writers are to be believed)
- they use a limited range of commands and few, if any, advanced searching features
- they are generally satisfied with their searching

- librarians are faster searchers and obtain more records
- there can be big differences in the searching behaviour and style of individual end-user groups.

2.3 End-user studies of practitioners

The information needs of faculty members and students are readily identified. They are, usually, working in one subject area, with specific problems to be solved, or hypotheses to be investigated and are generally highly motivated towards formal information acquisition. On the other hand, information-seeking in the media, government, industry, banking, law, etc. is a means to an end, whether that end be the production and distribution of wealth or serving the government and people. Formal information sources can take a back seat in these environments; speed of delivery and currency come especially to the fore; and information gathering can be a haphazard affair with many people unable to articulate their information needs, not knowing exactly what they want until they see it. Furthermore, with the full-text files that many practitioner end-users have available to them, librarians do not have the in-built searching edge, which their knowledge of bibliographical records and controlled language vocabularies gives them when searching bibliographic files in their own libraries.

A number of attempts have been made to understand the on-line searching behaviour of users in the non-academic world, and because of the more testing and less controlled environment they inhabit, provide a wider understanding of end-users. The early writing on the topic saw quite a polarisation of views. While some writers were convinced that managers, practitioners and researchers would appreciate being put in direct touch with on-line databases, and relish the opportunity to change ideas as the search progresses, without having to translate the query for an intermediary (Haygarth-Jackson, 1983), others were equally strong in their belief that end-users had neither the skill (Arundale and Erbach, 1988), time nor the inclination to learn how to search (Thompson, 1983; Quint, 1988). A disciple of the latter school, Charles Citroen (1984: 547) claimed that: *present day on-line systems often need complicated access procedures, require extensive training and continued study to be used efficiently... and pricing policies can induce very costly errors.* Some authors (e.g. Garman, 1989) felt that even if end-users took to on-line initially the novelty would soon wear off - and indeed in some places that is exactly what happened (Nicholas *et al.*, 1986).

Three classic studies carried out in the early 1980s shed the first real light on the problems that practitioner end-users in scientific and engineering environments encountered when on-line systems were first being introduced. A project at Raytheon (Richardson, 1981), a major US supplier of military sonar systems, was designed to test the hypothesis that end-user searching was more efficient that intermediary searching. After an in-depth training session engineers were given seven months' free searching time followed by five months at full rates. Despite the fact that the engineers claimed to comprehend fully the mechanics of searching, usage was astonishingly low (only 0.5 per cent of the free time available was taken up) And this low usage dropped dramatically during the full billing period (to approximately one-seventh of the previous low level of use). Why? By making access to information relatively easy the project designers had thought the system would be used. But engineers (in common with most practitioners) tend to be results-oriented, not information oriented; a new retrieval system, no matter what the physical form or no matter how accessible, would not in itself provide the incentive to seek more information. The study concluded: *From this experiment we would expect a high initial enthusiasm, a period of settling out, and a final cadre of dedicated, proficient users .... Librarians and information professionals themselves are not uniformly proficient online searchers, and it would seem reasonable that a similar range of skill will be observed among groups of scientists and engineers* (Ibid.: 48-49). It would have been interesting to learn what size cadre the author envisaged.

In the second study, Exxon (an oil company), gave all those interested - from the most senior chemists through to the secretaries, the opportunity to learn on-line searching techniques (Walton, 1983). Surveys carried out six weeks after the introductory training course and a year after that, again revealed low usage levels; even those who had expressed initial enthusiasm were not using the facility. Low use was attributed to the fact that: *online searching was a difficult and challenging skill, requiring time and effort to master* (Ibid.: 49). Users were discouraged by the careful preparation that proceeded the search and the logic involved in its execution. Also, the author pointed to what was then seen to be one of the intractable problems associated with end-user searching that: *even a well-motivated end-user with sophisticated searching skills will generally not have the need to search very often. And yet practice is necessary to maintain skills* (Ibid.: 49). Despite the lack of enthusiasm (and seemingly flying in the face of the their own evidence), the authors suggested that end-user searching should be encouraged and that all would come right in the end when *technology has advanced to the point that the entire search process is done automatically by a black box'* (Ibid.: 50). Thirteen years on and the black box is yet

to materialise, and yet end-use has grown dramatically in many fields.

The third study (Haines, 1982), again in the sci-tech field, but this time conducted at Kodak - appeared to contradict these findings. Over a twelve-month period, on-line searching by end-user chemists actually increased with the searchers reporting satisfaction with their self-conducted search sessions. It was on-line's time-saving qualities that were being appreciated. *Chemists are being paid to do research,* state the researchers, and *If an online searcher can accomplish in one hour what previously had taken a day or two using printed sources, one hour of online time by a chemist could save many weeks, even months, of time in the laboratory (Haines, 1982: 18).* This experiment supplied early evidence that, if given favourable conditions, end-user searching was feasible - but the conditions the authors were referring to at the time was a group well versed in literature searching and a literature that was well organised to boot (none of these conditions applied to journalists or politicians the groups at the centre of this study).

It would seem logical, that busy practitioners, deprived of the large document collections servicing their academic colleagues, would be more likely to respond favourably to information systems providing factual data and the full text of documents, than to bibliographic systems, which provide only signposts to potentially valuable items (and which, in many cases, they did not have readily to hand). Or so one might think. Certainly, the relative early successes of Datastream with the business community, FT PROFILE with journalists, and LEXIS with lawyers would lend support to this argument. And, of course, stock and commodity exchange services, like TOPIC, and wire services, like Reuters, have always been the monopoly of the end-user.

Encouraged by the early successes of these full-text services with end-users commentators began to write the epitaph of the intermediary. Jones (1984: 217) for one, thought that *the evolution of the professional online searcher must be regarded as a temporary aberration.* Perhaps this was just scare mongering, for there was very little research then (and now) to suggest that this would turn out to be the case. Charles Meadow (1983: 1622) was rather more optimistic about the future of the professional intermediary. For him, end-users would do their own more straightforward searches on a single familiar database, calling upon the information professional for the difficult more wide ranging ones. He painted the following, now famous, analogy: *Can I, the family consultant on cuts and bruises, do surgery? Do I need a lawyer for every dispute I may have with a shopkeeper? Of course not, I*

handle the easy, general uses, and the experts handle the difficult cases. This view increasingly found favour in later studies: and has stood the test of time. Other commentators consoled themselves with the thought that end users would soon tire: [there will be] *a large fall-off in end user searching after the initial novelty wears off* (Garman, 1989). Much of the end-user literature of the Eighties was written along these 'will they, won't they' lines.

With end-use eventually settling in and bedding down towards the late Eighties/early Nineties, we began to see a move away from studies primarily concerned with head-counting end-users to ones examining what part end-use played in meeting information needs and how effectively end-users searched (Nicholas *et al* 1991) - something which had ready happened in the academic sphere.

A few studies have looked more specifically at the differences between the searching patterns of practitioner end-users and intermediaries. Salovaara (1988) examined the differences in a technical research and development environment. Unfortunately, only fourteen out of a possible one hundred and forty researchers undertook searches in the survey period, recording only 48 searches between them. On the basis of this limited amount of searching it was concluded that: professional searchers found more relevant references per topic than end-users and this was largely achieved by the professionals searching more databases, though importantly from this studies' perspective, a half-professional searcher had as good a result as the professional and yet looked at fewer databases (this was interesting because MPs Research Assistants - a case study group for this research study - could be similarly classified); the goal of the professionals was high recall whereas that of the end-user was high precision - the latter achieved as a result of a simpler and tighter (more limited?) search strategy. The seeds of Lancaster *et al*'s (1994) thinking can certainly be found here.

With the introduction of CD-ROM in the early 1990s the end-user debate was given a new lease of life, and an opportunity to repeat the old hoary debates. It was argued now that CD-ROM, unlike on-line, was the true user-friendly search medium. Rigglesford (1992), for instance, cited its attractive layout, ease of searching and downloading (into spreadsheet packages for instance) and the possibility of graphic displays as CD-ROM's benefits, and thought that all this would result in a migration of use from on-line to CD-ROM in financial institutions in the City of London. He made the usual (pat) side-swipes at remote on-line systems complaining (on behalf of end-users) about the complexities of billing and communication frailties (an argument since shown to be short-lived by the rapid advance of the Internet). End-users, he

believed, would greet CD-ROM *with a source of great joy.* Yet currency above all else is prized by City economists - and this is something that - as yet - CDs do not score well on. The most interesting information contained in Rigglesford's paper was, that in a survey of users of his company's disc - FAME, it was discovered that end-users searched along fairly broad topical categories, whereas the searches of information professionals were typically more precise and sophisticated. Less surprisingly, perhaps, end-users tended to look for more depth of analysis of data and to make greater use of graphic presentation.

Rigglesford's claim for CD as an end-user breakthrough was no doubt influenced by the fact that he represented the company that sold the product he investigated. Certainly his enthusiastic claims are not shared by all. Thus Sullivan (1992), surveying the use of MEDLINE on CD by practising scientists, found that despite nearly 150 of them having been trained in its use, the system was only accessed 3 times a day; on average each of the 106 scientists who accessed the system used it once per month for about 22 minutes. Its hard to believe Medline hard-copy use was much less than this - certainly no revolution here. Plainly, training alone was not enough to guarantee high levels of end-use. Sullivan claimed also, that searching was done at a 'simple' level, but the data she presented does not seem to bare this out, for over half the search strategies contained between 2 - 5 words (a wide margin admittedly); nearly 15% contained more than five words; 30% of the strategies contained three or more steps; and that Limits were used in only 25% of the cases (an impressive statistic, surely?) The question that is begged, but not answered, was how did the librarians search?

The training - or more accurately - the lack of training of end-users has preoccupied many commentators. It is this literature, in fact, which yields the richest cache of end-user searching anecdotes and stereotypes. This librarian spoke for many when she said that: *I have frequently found people searching* [Medline CD-ROM] *in the most peculiar, inefficient and totally illogical fashion. It is a slightly hair-raising prospect when the searcher is a junior doctor sent by his chief to do the spade work for an article to be submitted for publication, let alone searches relating to treatment of actual patients* (Brewster, 1995). It was not just hospital doctors that come in for criticism. Barker (1993) singles out top research scientists for criticism. *Once you let inexperienced users do their own searching, quality can no longer be guaranteed. For instance, end-users may not appreciate the limitations of the data on the CD-ROM - that it only goes back to 1989 and may be up to three months out of date. They may not look at the thesaurus or the chemical index or think sufficiently about their search*

terminology - which means their search may not be thorough. We have also put **warning** [author's highlighting] *notices by the workstation and we are doing all we can to educate users*. It seems no group, no matter how eminent or well qualified, is spared the ire of some librarians.

Of course, CD ROMs are not the only user-friendly systems around (though not that user-friendly if the study above is to be believed), a popular dial-up variety is Easynet, a gateway system that provides standardised user-friendly access to 900 databases. Larsen (1988) in an end-user survey of Easynet, discovered that the menu-driven interfaces found in Easynet did reduce the number of no hits experienced by the end-user, but lots of no-hits there inevitably were. The author largely ascribes this to three factors: not understanding the nature and rules of on-line searching and its need for precision - they would for instance put in company names as the remembered them rather than in their official form; searching inappropriate databases; inability to transform information query into acceptable system search terms - users would enter such natural language phrases as *Mick Jagger and his life* (Ibid.: 471). Larsen also discovered that end-users experienced problems transforming an information problem into a search strategy and this most frequently manifested itself in the broadening of the query: thus one user asked for information on new cars when they actually wanted dealers prices for the Honda Civic. (Ibid.: 472)

A number of writers do not think that user-friendliness is the real obstacle or issue though. Koblas (1993), for instance, opines that, contrary to perceived wisdom, accessibility and ease of use are not prime determinants in people's use of information systems. She detects quality (of the data) considerations too and believes that it is not so much ease of use that is important but perceptions of how easy it is to use. For her, it is even more complex than that, for she would add personality, environmental and social characteristics as important determinants as to whether someone goes on-line or not. Of course, these are also general information seeking determinants. Oppenheim (1992) is of a similar mind, for as far as he is concerned Hosts too readily believe that their failure to crack the end-user market is because their product is technically lacking: their solution - as FT PROFILE's is - is user friendliness: menus, gateways, Windows, CD-ROM etc. According to Oppenheim, this emphasis is misplaced, for what really concerns the user, is whether the system offers *the answer to the person's questions - the right information at the right time at the right place*. In defence he points out that spreadsheet programs sell hundreds of thousands of copies despite their well-known complexity. This of course runs counter to conventional wisdom in the information field, that by minimising user effort through improving ease of use, the

result will be increased use of an information system.

Despite the fact that many end user groups have now been active for a good number of years now, the indications are that there has not been much of an improvement in searching skill, although, admittedly, the evidence is thin. Fisher and Bjorner (1994: 282), recently questioned a large number of American librarians about their end users (lawyers amongst them). Many expressed the opinion that end users conducted simple searches, were not up-to-date on their techniques, did insufficient searching to maintain the skills they obtained in training, and that it was acknowledged that *the library can answer questions faster, cheaper and in greater detail* - the last point of course underlining the librarians concern with recall. Miles (1993), reviewed in the following section, also provides support for the view that little has changed; indeed, in her case things might well have regressed.

3.4 End-user studies of journalists and politicians.

Journalists

It seems obvious that, where users are already involved in investigation and research, where their information seeking is sometimes cloaked in secrecy; and where they have access to full-text databases and downloading facilities, there is a great incentive to search themselves: and all of this very much applies to journalists. And as mentioned earlier, with full-text files librarians do not have a natural advantage over journalists, which their knowledge of bibliographical records and controlled language vocabularies gives them when searching bibliographic files.

The published literature on journalists as end-users is insubstantial when compared to the enormous literature on academic end-users. Much of the explanation lies with the difficulty in obtaining the co-operation of a busy, preoccupied, unpredictable and mobile group of people. They do not lend themselves to the type of laboratory examinations that academics appear to willingly subject themselves. It cannot really be anything else because, journalists offer researchers an interesting, high profile and strategic group of information users to study.

One of the very first studies conducted in this country on journalists as end users was that by Harman (1986). In a questionnaire study of Reuters' journalists use of their own in-house database, Newsbank - a recently introduced substitute for hard copy cuttings - it was found that: (1) only 3% of journalists used it more than five times a

week, with more than 50% saying that they never used it; (2) no one found Newsbank very useful, and only 15% thought it useful; (3) three-quarters of journalists had little or no understanding of it. (explaining the above statistic perhaps). Journalists also felt that they were insufficiently trained in its use and that supporting documentation was poor or missing (50% mentioned this). Others wanted the library staff to search on their behalf and cited the poor currency of Newsbank as a major problem.

Questions were also asked as to the journalists' use of NEXIS, a full-text current affairs service to which journalists had access for more than a year. The answers to these questions showed that NEXIS was used more than Newsbank, that 6% used it more than five times a week; just under half used it 1 - 4 times a week; and a third not at all. One of the principal reasons for the greater use of NEXIS was that it was easier to use, though probably not many people outside Reuters would recognise it as being such. The general tenor of the survey was that journalists would have preferred to have librarians do their searching for them, especially so when the system was perceived as being difficult to use. Harman concluded that, journalists saw *the concept of a journalist-used - i.e. end-user - database as a waste of time, taking the journalist away from the main professional functions of writing or sub-editing.* Jennifer Harman's study of Reuters demonstrated that a good number of journalists had it in them to become regular end-users, but also provided early warning that the road to end-use was a rocky one.

A BLRDD funded project Information Seeking in an Information Society shed light on the on-line habits of journalists post Wapping (Nicholas *et al*, 1987; 1988). A confused world was described, with some journalists taking to on-line with alacrity (chiefly those from *The Guardian* and News International), others fearful of its influence and some even banned from using it (those at *The Independent*). Some of the raw data collected from the research project can be found in the chapter on journalists, and the results of the study are discussed again in the review of work conducted at News International.

Disregarding Harman's findings, Arundale and Erbach (1988: 577) made the standard claims that journalists did not possess the necessary skills to search multi-database hosts and believed that even journalists would concede that *the library's staff ... are more likely to be successful with [the] more complicated or sophisticated search* (though none are quoted as saying so). They also inferred that, because library staff search constantly, that made them better searchers and that journalists would be unwilling to train to become proficient searchers (that assessment, at least, has turned

out to be true). No data were presented to support these statements, and it is probably significant that the authors came from a newspaper where end-user searching facilities were not provided as a matter of policy.

Jacobson and Ullman (1989), in examining the impact and use of commercial databases on US newspapers, conducted a large-scale questionnaire survey of 80 journalists. They found that journalists used databases - mainly - VU/TEST, NEXIS and DIALOG - *quite frequently*: 44% once a week, and 23% more frequently than that. Home (Metro) and Business where the departments that used on-line the most and Sports and sub-editors the least. Journalists were generally happy with their searches - three quarters (76%) said their searches were *almost always* useful. Few experienced any problems in searching - expense and technical problems were the only handicaps really mentioned. The value of these responses were somewhat undermined by the fact that the journalists were not directed to a specific Host when they answered these questions - one would have been especially interested in their views on DIALOG, an allegedly difficult system for journalists to search (Jackson *in* Nicholas *et al*, 1987). Journalists largely used on-line for background information, though fact-checking, preparing for interviews and assessing national/local trends were also mentioned.

Two studies have been conducted of News International journalists, where the biggest concentration of journalist end-users in the UK can be found. One of these studies (reported in Nicholas *et al,* 1987) was conducted shortly after they moved to their high technology environment at Wapping and, the other (Miles, 1993), five years after that. The first was a questionnaire-based survey that targeted both broadsheet and tabloid journalists. Forty-eight broadsheet journalist responded to the study, which, interestingly, was instigated by the journalists themselves - worried about management cuts in their burgeoning on-line searching. Over three quarters of the journalists reported that they *frequently* used on-line databases (in fact just, FT PROFILE): this meant that one in three journalists at the newspaper used PROFILE *frequently* (what *frequently* meant was not actually defined). Only one journalist had not used the service. Foreign, business and the investigative teams were the most active users and features, sport and sub-editors the least - very much as Jacobson and Ullman (1989) had found. Given the siege mentality prevailing at Wapping at the time, these relatively very high rates of on-line use can easily be understood - it was the journalists' information life-line. Management, however, attributed the alacrity with which journalists took to on-line to laziness (*too lazy to walk a hundred metres to the library*) and to their *spend* mentality (*it's in their nature to be big spenders ... they*

37

book the best hotels, travel first class and run up big online bills (Nicholas *et al,* 1988: 23). It was not so much that the volume of end-use that was impressive, but the passions aroused by its potential removal: *Datasolve* (FT PROFILE) *is the best thing about the place, journalistically speaking; The single most useful* (information) *tool we have; Datasolve has become absolutely indispensable* (Ibid.: 24-25). High praise indeed.

In the comments liberally spread around the questionnaire, journalists proved particularly perceptive. They seemed to readily understand: the attractions of on-line as compared to the journalist's traditional information store - the cuttings collection; on-line's interactive and cross referencing facilities, the speed with which information could be delivered on-line, and on-line's special attraction to a new and still experienced staff. Problems were also recognised - as were the skills necessary to overcome them: *once you have mastered its ability to trawl through masses of material using the headline, context and text search facilities it's unmatchable* (Nicholas *et al,* 1988: 25). Fascinatingly, this quote also showed how important display commands (headline etc.) were to journalists in searching and refining their searches. Journalist end-users appeared to lump search and display commands together. Lack of currency, limited source coverage and gaps in the allegedly full-text coverage of individual newspapers (Sport was seldom included then, nor were the arts columns or letters) were other problems mentioned. Many of these problems have since been rectified.

The tabloid journalists in the survey - there were 57 of them - were not such heavy users of PROFILE, with just a quarter of them using the system regularly and quarter not using it at all. However, these figures have to be taken in context, for at that time tabloid journalists at Wapping, unlike their broadsheet colleagues, did not have their own newspaper available on PROFILE (they are still not on). This might well explain why they were more willing to express their dissatisfaction with the system: complaints largely involved technical problems (the system hung easily in those days), though a number found it too complicated and slow to search. The conventional wisdom at the time (and now, to some extent) was that tabloid journalists would only ever search a database if the facsimiles of the stories were present. This survey proved that this need not be the case. There still appears to be some mileage in the belief though (Poynder, 1993).

More than five years on, and if anything, things at Wapping seem to have slipped back somewhat. In a (very) small-scale dissertation study of end-users at *The Times*

conducted by Miles (1993), involving just six journalists (for this reason the findings need to be treated with a high degree of caution), it was found that four of these journalists used FT PROFILE, although none were trained in its use. (This, seemingly, being par for most end-users). Furthermore, two journalists did not even have quick-reference sheets to assist them: they managed by asking colleagues for help/guidance. Despite the alleged user-friendliness of the system, three journalists thought this to be a handicap. On-line help was not called upon, even though some of the journalists were aware of its existence. Indeed, journalists expressed irritation with on-line notices - something also noted by Nicholas *et al* (1987: 78). This suggests that menus are not the answer to journalists' on-line problems. Interestingly, none of the journalists thought it fit to obtain help from the Library, despite the fact that there were accomplished on-line searchers there. (By way of explanation journalists at Wapping work quite a distance from the library.)

On average, journalists searched the system twice a day. The typical use to which the system was put was very similar to that described by Jacobson and Ullman (1989): facts, figures, background research, previous treatments of stories and assistance in finding people for interview (quite specific uses of an on-line system). A number of file groups were used to find the information - but no more detail than that was provided. Most of the journalists thought the system easy to use, though it transpired that this was because they searched extremely simply - only one journalist availing himself of the Boolean operators or word proximity devices: **get** and **pick** seemed sufficient for most. Yet, despite broad and crude searching, and the inevitable fruits of this on a full-text, free-text system (lots of noise), nobody complained about receiving too much data - journalists generally don't (Kelly, 1995). As ever, what loomed most as the problems were the practical and physical side of things - logging on, scrolling and downloading data. Only one journalist saw the Library as a viable alternative to conducting the search themselves - all the others preferred to exercise control over their own searching, even if this meant waiting to go on-line.

In something of a turn around, Stanbridge (1992), a journalist, conducted an interview survey of UK media librarians, to discover facts and attitudes in respect to journalists' searching databases themselves. He pointed to the very low-rate of take-up of on-line by broadcasting journalists - something also chronicled by Nicholas, *et al.* (1988), and puts this down to ignorance of what was available and the prohibitive cost of searching on-line databases. What he failed to mention was that broadcast journalists typically have researchers to help them - and it is to them that they generally delegate the formal information seeking task. On the basis of anecdotal evidence supplied by

librarians (some, it seemed, with an axe to grind) he concluded that, even where journalists had access to on-line databases, they searched them poorly. Typical of his colourful descriptions of how journalists searched is this one: ... *reporters working under pressure usually bait their hook with a random keyword or two and are quite happy with anything they catch that adds some colour or gives a slant to what they are writing* (Stanbridge, 1992: 47). Vreekamp (1995), was another journalist with something to say about his colleagues and their use of on-line databases. Reporting more recently than Stanbridge, and concerning himself much more with Continental Europe, he found: that no correlation existed between level of education and intensive use of on-line databases (a strange line of enquiry to take, although no data was presented to support this assertion, however); a strong preference for the traditional (oral) means of acquiring information - on-line databases being relegated to the role of fact-checker; and that *journalists had uncertain feelings towards ... electronic storage* (Vreekamp 1995: 49). Given the paucity of data provided - and its anecdotal nature - it is difficult to judge whether Continental European journalists are really different to their colleagues in the UK in their use of on-line databases.

In the US, where journalists were introduced to databases a few years earlier than their UK counterparts, Ernest Perez (1980) confidently predicted some fifteen years ago that o*nline news library systems will give new powers to editorial staff, permitting a new kind of journalistic investigative approach.* However, Leonard (1992: 1), writing some 12 years later, was still promoting the virtues of computer-assisted journalism, claiming that *journalists can get the computer to do the same work* [as] *those dogged veterans did manually.* Some US schools of journalism are now actively teaching students the necessary techniques, so perhaps, the potential will soon be realised. Certainly the opportunities must be there, but in the UK anyway, there is scant evidence that the opportunities are being taken. Elsewhere in his paper, Perez (1980: 22) quotes Philadelphia Newspapers' Joe Di Marino as saying that *self-service on newsroom full-text system VDTs have quickly resulted in the elimination of about 80 or 90 per cent of the information requests coming into his library.* A somewhat threatening statistic this, as far as media librarians were concerned, but as yet (15 years on) there is no sign of this happening in the UK press, though broadcast libraries have been devastated, not so much by technological factors as by economic and political ones (Nicholas and Pandit, 1994).

A recurrent theme of much of the on-line literature on journalists, on which we have already briefly touched, is that journalists will prove resistant to searching themselves because hosts, like FT PROFILE, supply just text and not the image - and the image

is allegedly very important to them. Thus Poynder (1993: 8), quotes a News International librarian as saying that, *to journalists the value of a newspaper story is as much in seeing where it appeared on the page and what it looked like, as [for] the information it contained. ... where a story appears on the page indicates how important it was*. This obviously applies more to tabloid journalists than broadsheet journalists (but even so we have seen earlier that tabloid journalists will use PROFILE), and this view has been the principal driving force in the move towards introducing costly optical storage systems in newspaper libraries. Significantly, no broadsheet has yet such a system up and running, though News International, with their mixed broadsheet and tabloid stable, are heading that way. Nor, has the lack of an image prevented end-use reaching dizzying heights at Wapping - costing the paper half a million pounds a year, and bringing forth the claim from the same librarian, that *News International now spends more than any other Fleet Street paper on end-user access and we expect that growth to continue into the future* (Ibid.: 8).

Journalists' on-line searching has seldom been considered in the light of their wider information seeking behaviour. But this is true of the general end-user literature, too. A rare exception is the study by Kelly (1995). Kelly in a broad examination of information needs and on-line searching behaviour, conducted interviews with a dozen feature writers, most, but not all, from the tabloid press - *The Guardian, Spectator* and *Times* were also represented. In respect to on-line searching the key findings were that: (1) five had access to on-line services - and another one was not sure, but none of the journalists were able to conduct their own searches; only one could mention FT PROFILE by name; four had librarians conduct on-line searches on their behalf - once a month being the typical frequency of searching; a favoured (delegated) request was a search for articles by themselves; journalists were ignorant of the capabilities of on-line; cuttings were preferred, partly because of their browsing characteristics and partly because they faithfully reproduced the image - support for Poynder's(1993) thesis here. Some of Kelly's more general remarks are of value in understanding how journalists search: (1) their information appetites seemed insatiable - and the author believed that there lay the future attraction of the Internet (but why the Internet and not FT PROFILE one might ask - maybe, hype or fashion has got something to do with it?); (2) they are nervous about intermediaries filtering information on their behalf (but why then do they not do their own on-line searching?); (3) they are big browsers, never certain of what they want or what they are going to find.

Politicians

If the literature on journalists as end-users is sketchy, then that on politicians is almost non-existent. There have been no substantial studies of politicians and their use of on-line, published either in the UK or US. This is due in part to the fact that politicians readily delegate all information seeking, especially that associated with information technology, but it also points to the general neglect of the information seeking behaviour of this strategic group of social science practitioners. Things are, perhaps, about to change for The House of Commons began the piloting a Parliamentary Data and Video Network (PDVN) in 1993, which involved 105 Members and their Assistants. The provision of a PDVN has been mooted for many years now (House of Commons Services Committee, 1984; 1991) seems now to becoming a reality. Providing MPs and their assistants direct access to POLIS, CD-ROM networks and commercial on-line services is very much at the heart of this development. A report on a pilot study, provides some interesting data on end-use (House of Commons Information Committee, 1994) and is discussed more fully in the Postscript section of the Politician's case study. There were three key sources of data in the report: (1) a questionnaire study canvassing Members opinions on what they would like from the Network - this obtained 327 replies; (2) a questionnaire study of members and their staff forming the pilot group; (3) the submissions of MPs and expert/interested organisations and departments to the Committee.

The large-scale questionnaire study found that, although most MPs had IT facilities in their offices, it was largely their untrained staff who used them. In regard to on-line searching: 20% claimed to have used POLIS; 50 Members had direct office access to POLIS from their offices, 6% used Hansard on CD-ROM, and 5% FT PROFILE. As to the to the future - courtesy of the network, half were looking forward to using POLIS and Hansard on-line, 35% the Press Association newswire (made available on PROFILE in 1995) and 31% FT PROFILE. The mention of the last two demonstrating clearly MPs interest in current news and affairs.

A questionnaire study of pilot PDVN users - 34 MPs and their Assistants, showed that, POLIS was used most frequently, the networked CD-ROMs next, with FT PROFILE trailing in third place. However, of the IT facilities used, FAX was thought to be most *useful* (11 respondents thought so): demonstrating how high speed of delivery figures in the information needs of politicians. CD-ROM - Hansard and daily newspapers - was a close second (10). Despite being frequently used, POLIS was not thought to be most useful IT facility - only 6 Members though so (bibliographic systems rarely captivate end-users); more surprisingly, FT PROFILE was though

most useful by only 4 MPs. There were complaints about on-line systems, too: CD-ROM was thought to be *user hostile, so out of date it is useless;* POLIS fell down on the fact that it did not have a *full text facility,* it was *user hostile,* and *very slow;* and in the case of FT PROFILE *dialling out was difficult.* (Ibid. 112) - a problem that MPs seem to share with journalists.

Access to FT PROFILE was provided on a short term (12 months), experimental basis. It was not said so in the report but this must have been a ploy on the part of PROFILE to gain a foothold in what could prove to be a lucrative market. 13 members were signed up for the experiment and given training and a search allocation, but few Members took up their allocation, leading the Library to the opinion that *most Members are content for Library staff to carry out searches on their behalf* (Ibid., 99). Quite a difference to journalists, then, who would have fought tooth and nail to retain access to the facility.

The corresponding views of the Research Assistant on The Parliamentary Data and Video Network was recently provided by Watson (1994). The author points to the problems of overdosing on information as a result of MPs and their assistants being connected to data networks. Politicians (and journalists for that matter) have huge mailbags. With the introduction of the PDVN and the resultant exposure of MPs to IT, we can probably expect a surge in research studies. In fact one has already begun (Martell, 1995).

Ironically, what appears to be happening now is that the PDVN is in danger of being eclipsed in MPs' minds by the Internet: proving MPs are, if nothing else, fashionable. As usual when it comes to information matters the Opposition is making the running: *The Independent* reports (Victor, 1995) that Labour outperforms the Conservatives by three to two in connections. But significantly, both parties seem very bad at using it, though this has not stopped all three parties using the Internet for their party 1995 annual conferences.

CHAPTER 3
METHODS

3.1 Methods used to evaluate end-user searching

When it comes to evaluating end-use there are an extensive range of methods on offer. There are user surveys (questionnaires and interviews), focus groups, observation, talk-aloud techniques, precision/recall tests and, the one featured increasingly more and more in published research, transactional log analysis. It was simply not just a case of choosing the best one from this list, because each method has its strong and weak points; equally, each method highlights certain aspects of end user searching and neglects others. No end-user study could claim to be comprehensive unless both quantitative and qualitative research methods were employed. Hancock-Beaulieu (1989B) points out that, the data gathering method determines the type and quality of the data collected and to obtain a truly comprehensive or holistic picture of searching behaviour it is necessary to adopt as wide a range of these methods as proves feasible. Hancock-Beaulieu also makes a distinction between systems-oriented methods and user-oriented ones. Given the strong user need focus of this study it was inevitable that the selection involved the latter group. In consequence, such, measures of system performance as recall/precision tests and prototyping were rejected at the outset.

Inevitably, too, the choice of methods is determined not just by the issue being studied (end-use), but also by the user group being investigated. Thus there was little chance of getting busy and self-important practitioners, like journalists, politicians, lawyers etc. to subject themselves to the kind of examination that information researchers subject students to - talk-aloud methods and on-line questionnaires were the methods the researcher has in mind. Also, academics are generally a captive audience, whereas politicians and journalists are certainly not - and this has a big impact on the conduct of the study.

Another factor in the choice of research methods was the amount of time that the researcher had available to give to the study. Conducting research with such large groups of people, over such a long-period of time, at two geographically remote and separate locations, inevitably meant that methods had to be chosen with these constraints in mind. Therefore, methods offering remote and automatic monitoring (transaction log analysis), or that were self-administering (questionnaires), proved attractive. Other methods that might have otherwise appealed, like before and after

(search) interviews, had to be rejected because of the demands they would have made on the researcher's (and user's) time.

As a result of all these factors, the number of methods were whittled down to four: interviews, questionnaires, observation - though the latter only figured briefly, and transactional log analysis (for the sake of brevity, termed logs from now on). All these methods featured in both case studies. The precise combination of methods was made with the object of obtaining as broad and detailed a picture of searching as possible; indeed, the attraction of the case studies were that they both offered the opportunity to examine end users in a wide and rich variety of ways, and in the context of their information needs. The fact that practitioners were being studied inevitably meant, that a significant number would not be end-users (you couldn't imagine students absenting themselves from OPAC searching in the same numbers): therefore methods monitoring use alone (e.g. logs) would not be enough in themselves. Social survey methods have an important role to play here. The actual deployment of the methods was dictated by a number of factors, not all within the researcher's control. Thus logs did not become available for *Guardian* journalists until right at the end of the study. Some methods were more appropriate at certain stages of the study: thus observation and interview were especially useful in the early stages of the project when end-use was in its infancy. In the particular case of the *Guardian*, where research stretched back more than ten years, on-line was at such an early stage that there was little choice but to choose observation and interview as the research instruments.

The very nature of the study - studying two different groups of people who were using different on-line systems, in different locations and at different times - inevitably creates problems when presenting an integrated, rational description of the methods used. And that is why each of the case studies have been provided with its own methods section and conclusions. Nevertheless, if confident comparisons are to be made between the two sets of data and the results generalised then it is incumbent upon us to do this. Tables 3.1 and 3.2 provide a summary and comparison of the various methods used. From Table 3.1 it can be seen that, because of the need to capture the whole searching process, the same mix of methods has been used in both studies. Furthermore, in both cases the various methods used fulfilled the same roles - computer logs to obtain the fine detail about the time spent on-line; interviews; questionnaires and observation to obtain: (1) the before and after, (2) an explanation for the behaviour on-line; (3) the attitudes of the users, (4) the views of non-users. The differences emerge at the more detailed level. Thus sample sizes differed - partly

because of the difference in size of the two populations being surveyed and partly because it was recognised that co-operation would be more forthcoming at *The Guardian,* because of the support of the library staff and the close proximity of the users. The number of respondents differed too, reflecting different levels of co-operation and good-will; as did the logs, which differed in the amount of detail captured and the time period monitored - this being a function of the differences in the monitoring software and Host co-operation. The topics covered by the questionnaires and the phrasing of the questions, although generally similar, differed in some details - partly because the questionnaires were distributed at different times (four years apart), partly because of the different issues that concerned each group at the time and partly because they were the result of collaboration between researcher and the library staff involved.

In an ideal world user surveys and log analysis would have been undertaken at the same time - and, indeed, this is what largely happened at The House of Commons. In the case of *The Guardian* two factors prevented this happening: (1) the necessary co-operation from the relevant commercial host (FT PROFILE) was a long time coming; (2) resources simply would not allow it. Also, *The Guardian* study was conducted over a long period of time, and it was very doubtful whether the co-operation and will would have been there to have sustained a single research method over the projects' duration. Of course, when a variety of methods have been used over a relatively long time frame it is sometimes difficult to know to what to attribute changes over time to - to real change in user behaviour or changes in method? As it turned out this was not so much a problem, for in most cases they told largely the same story.

While these differences mean that caution has to be exercised in making comparisons and forming generalisations, the researcher is of the opinion that this is the price that has to be paid once you attempt to conduct wide-ranging end-user studies outside the largely controlled academic environment. In mitigation perhaps, the study never set out with the idea of examining end-use from a laboratory or systems point of view, it was always envisaged that end-use should be examined in the dynamic and complex environment of the workplace.

Although the two case studies employed the same mix of methods and were subject to a very similar framework of analysis, they were conducted over different time periods (1985-1994 in the case of *The Guardian*; 1989-1991 in the case of The House of Commons). Partly because of this, partly because the on-line systems under investigation were so different in character, and partly because the methods impinge

so closely on the results of the studies, it has been thought best to place a detailed explanation of the methods associated with each case study with the case study itself. Here instead will be found a general account of the methods and their attractions and an examination of how some key research studies have employed the methods in their studies. Given the large and central role transactional log analysis plays in both studies and its relative novelty in studies of practitioners, it has merited rather more in the way of explanation and discussion.

TABLE 3.1
A SUMMARY AND COMPARISON OF THE METHODS
USED AND SIZE OF POPULATIONS STUDIED

Location	User group	Method	Sample size	Respondents (Survey) Users (logs)	Sessions captured
House of Commons	End-users	Questionnaires and interviews	115 RAs 10 MPs	60 (52%) 15 interviewed 10	n.a. n.a.
	(n=600 RAs)	Observation	n.a.	10 MPs; 12 RAs[3]	n.a.
	(n=650 MPs)	Logs	1 year	110-120 RAs[4]	1307
	Library staff	Questionnaires	-	-	n.a.
	(n=59[1])	Interviews	22	22	n.a.
		Observation	n.a.	3	n.a.
		Logs	1 year	59	11161
The Guardian	End-users	Questionnaires	120	87 (73%)	n.a.
	(n=240)	Interviews	14[2]	18	n.a.
		Observation	n.a	9	n.a.
		Logs	2 days	39+	86
	Library staff	Questionnaires	12	8	n.a.
	(n=12)	Interviews	8	8	n.a.
		Observation	-	-	n.a.
		Logs	2 days	8	53

1. Just counting staff who searched on-line for users. 2. Includes 3 repeat interviews with one user. 3. Research Assistants 4. Estimate

TABLE 3.2
A COMPARISON OF THE SUBJECT COVERAGE
OF LOGS AND QUESTIONNAIRES

	LOGS		QUESTIONNAIRES	
	House of Commons	*The Guardian*	House of Commons	*The Guardian*
Frequency of use			√	√
Session numbers	√	√		
Search numbers	√	√		
Transaction numbers	√			
Documents displayed	√	√		
Zero hits	√	√		
Screens viewed		√		
Connect time	√	√		
Duration of search	√	√	√	√
File selection	√	√		√
Retrospective searching	√	√		√
Time of searching	√	√		√
Interface preference			√	
Type of search	√	√	√	√
Fields searched	√	√	√¹	√
Search commands used	√	√		√
Search steps		√		
Terms used	√¹	√		
Display commands used		√		
Errors made		√		
Success/satisfaction	√	√¹	√	√
Training			√	√
Problems			√	√
Access			√	√
Delegation			√	√
How found out about			√	

1 Indirectly only

3.2 Transactional log analysis

Logs were used to chart end-use at The House of Commons and at *The Guardian*. They were however different in the detail they supplied (partly because of the different systems involved and partly because of the different software doing the monitoring), the duration covered (POLIS's logs spanned a year, and *The Guardian*'s two days) and the numbers of users they monitored. See Tables 3.1 and 3.2 for more details.

According to Hancock-Beaulieu (1990: 523) transaction log analysis is the automatic monitoring activity of the computer system. Peters *et al* (1993: 37) add flesh to this definition: *transaction log analysis can be narrowly defined as the study of electronically recorded interactions between online information retrieval systems and*

the persons who search for the information found in those systems. Helpfully, the authors also define the most problematic aspect of it all - what actually constitutes a transaction - and we are told that this is the users' query and the systems response (ibid.: 39). Despite the fact that transactional log analysis has been with us since the late Sixties - introduced to study systems' performance (according to Peters *et al*, 1993), it was only really since the Eighties, when the widespread introduction of OPACs and CD-ROMS into educational institutions gave researchers lots of motives and opportunities to study end-user searching, that it became really popular. Logs record certain things, typically: the characters pressed by the user before a carriage return is initiated; the time, date and location/source of transaction; the system's search command identifier and the systems response, such as screens displayed or items located. However, different information retrieval systems measure different things - OPACs will undoubtedly have different logs from those of full-text systems for instance. In this study the POLIS and PROFILE logs recorded different things. Though most offer the opportunity to measure: motor skills (typing); basic knowledge/level of education (spelling); conceptual knowledge of searching as well as the use of the system.

Peters *et al* (1993) identified a number of key strands to log analysis research: (1) the use of the full range of system commands or advanced searching features; (2) the number of characters or terms input by users, and whether the terms matched the words on the database; (3) response times. In truth there are many more strands than that, for so much can be read into logs. In fact there lies one of the chief problems with logs - the temptation is to read too much into them. This can lead to the production of research that is sometimes of dubious worth or contradictory in nature. This is particularly evident in the work done on the topic of zero hits. On this aspect of searching researchers have had a field day (Zink, 1991; Peters, 1989). However, Larson (1991) believes that there are so many variables associated with on-line searching that a wide variation in results is almost inevitable: system differences, user differences, variations in data collection methods - even slight ones, and differences in the time period covered can lead to quite different results. This has to be borne in mind when interpreting the results of this study, for all these factors are present in it.

The attraction of transactional log analysis as a research instrument in the area of end-use is as follows :
1. Not a lot of labour or cost is involved, and as such it is a great improvement on labour intensive methods like interviewing. Having said this though, while the actual collection of the data is sometimes simplicity itself - tempting the researcher

to bite off more than they can chew, the subsequent analysis can be long and painstaking, especially if the researcher is looking for something more than a straight numerical count.

2. To a profession generally starved of hard statistical data, the sheer volume and level of detail that can be captured is beguiling. Surveys and focus group sessions can never hope to match logs in this. Questionnaires *et al* too often provide generalities, logs provide the detail. Interviews and questionnaires can alert and point the researcher to user's search styles, but too often they provide insufficiently detailed information to incorporate this into information system design. To know that someone uses, say, truncation regularly is one thing, but what does regularly actually mean? There is also a need to know exactly how many times it was used, over a period of time. To be able to describe searching characteristics in terms of hundreds or thousands of incidences, rather than dozens of them, gives log studies a certain weight and authority - not always deserved, of course.

3. The objectivity of the method is one of its strong points: users' attitudes towards librarians and the system does not effect the result. Research findings are too easily biased by the fact that: academic and public library users are often supplicants in the exchange of information, not wishing to be seen making comments that might appear unduly critical; telling librarians what they think they would like to hear.

4. The method is unobtrusive: no one is likely to refuse to take part in the study - largely because they do not know that it is going on, so there are no problems of low responses or biased samples. However, there are some big ethical questions to consider: do you inform all users as to what is happening and what happens if some refuse to co-operate and your system does not allow you to identify individual users? Also logs, shaved of individual identifiers, cannot be explained in the broader information needs context. Fortunately in one case study (*The Guardian*) it was possible to identify searching to actual named individuals and in the other (The House of Commons) you could frequently pin down searches to small groups of user.

5. Finally, logs tells it as it was, not, as in the case with interviews and questionnaires, as what users actually remembered.

However, transactional log analysis is not the panacea that is at first appears to be. Problems do arise:

1. Because of the superficial nature of the data: a problem common to all forms of statistical data, bibliometric analyses in particular. Yes, lots of numbers are there,

but what they all mean is not always certain. Logs simply plot a particular form of behaviour, and even then, it is not always possible to point to a motive or reason for a particular search operation. Additionally, logs do not record users' needs or intentions, nor do they measure their satisfaction: to establish these things we must interview or send out questionnaires. There is a need to capture more of the search process - the before and after for instance, and it is to social surveys that we look to obtain attitudes, opinions and perceptions of success. Lancaster *et al.* (1994) maintain that the real interest for researchers lies not in the fact that a command was used x times, but why certain approaches were adopted and others rejected, and, as a result, prefer to see users talking through their actions at the terminal.

2. In assigning data to an individual user or even a category of user, especially in the case of on-line open-access stations, where logging on and off is rarely undertaken. Hancock-Beaulieu (1990b) overcame this problem by observing the sessions and noting down the start and completion of sessions. This was a problem with the POLIS in-house searching, but given the duration of the monitoring (12 months) it would not have proved feasible to adopt Hancock-Beaulieu's approach. Even when users have their own on-line facilities, other people may use them from time to time, when they are on holiday for instance. A related problem occurs in the case of monitoring intermediary searching - and, that is distinguishing between housekeeping use of the database and genuine searching on users' behalves. This problem had to be dealt with in the case of POLIS use at The House of Commons.

3. When interpreting the raw data. Peters (1988) highlights the dangers of reading too much into logs. She explains that just because logs show users conducting high recall searches, for instance, it does not necessarily mean that they prefer such searches - this may simply be a function of system design or untutored searching. Take also that concept of zero hits. Peters claims that a zero hit search does not necessarily mean that the user had a problem with the system or that the search was a failure - it might have simply been a case of the library not having what the user wanted. (the search, for instance, might have been a particularly good one).

4. Logs are generally full of surprises - gaps, unexpected occurrences, the inexplicable. Somehow, if there is something unusual or odd or unknown about searching, logs will magnify the problem, but not usually provide sufficient insight to explain it. Thus for instance as far as HOC library staff were concerned no Research Assistants used the command-driven version of POLIS - the logs showed that some did.

5. They can produce contradictory results. An example: Bennett (1975) saw no relationship between the frequency of terms used in the indexing of the database

and the frequency of terms input by users, Nelson (1988), however, found what appeared to be the exact opposite - that users often searched on (title) words that were found in large numbers on the database.

A number of aspects of on-line searching behaviour can be gleaned from logs, and these are examined below. What follows applies not solely to logs, for many of these aspects can be studied through other methods, but they are most clearly and dramatically seen in the analysis of logs, and thus are discussed here in the most detail. The various aspects of the search captured by the logs can alert us to different things. Indeed, a single element, like the duration of a search session, can tell us a number of things about the user: the fact that they are a fast/slow searcher, efficient/inefficient searcher; or active/inactive searcher. Also logs enable a direct and detailed head-to-head comparison of the searching behaviour of end-users and information professionals - and as this is a key aspect of the present research it is not so surprising that they commend themselves so highly. The fact that there are differences in behaviour between the two groups can imbue even a seemingly minor aspect of on-line behaviour with interest. Comparisons can be difficult though, when, as was the case with POLIS use, librarians and end-users search the system in different modes - respectively, command and menu-driven modes.

It is probably most useful to view log details in terms of what they can tell us about facets of on-line information seeking behaviour: the volume and frequency of use; characteristics of searches; problems encountered; and levels of success and satisfaction. These topics are not mutually exclusive and there lies another problem associated with log analyses.

Measures of use or activity

Obviously one of the major points of interest in all on-line evaluations is whether the system is used: how much the system is used and whether use varies much from group to group and especially whether, as alleged, end-users are generally less active than Library staff. As the literature review has shown, use and non-use is at the very heart of most end-user investigations. However, use can be measured in a variety of ways, and because some of these measures are distinctly problematical, it is always best to employ a range of them in any investigation. The logs examined as part of this study offered the following indicators of use: (a) search sessions conducted; (b) searches conducted; (c) on-line transactions made during a result of a search; (d) the length of time spent on-line; and, (e) records displayed and screens viewed. On-line host are

interested in use indicators, too, because generally that is how they charge for their services. By their charging systems hosts recognise that, as far as they are concerned, there are just two key components to use: the latter two - connect time and data (lines or records) displayed.

Number of on-line search sessions conducted

What constitutes an on-line search session is not obvious and hence there is a need to define the term. Simply, it might be defined as the on-line activity that occurs as a result of a particular query or line of enquiry. Practically it is identified by a log on and a log off, although it is always possible for a user to log off and return to a search later - maybe after a period of refection (there was some evidence of this in the logs of *Guardian* librarians). In connection with in-house information systems, like POLIS, there is a problem of identification and demarcation, for these systems tend to be left on during the working day: thus users do not always mark their arrival or departure by logging in or logging out - many sessions are not formally terminated, they are just left in the air - and picked up by the next person. With commercial, connect time payment systems, like FT PROFILE, searching is inevitably much more disciplined, and as a consequence a lot more can be read into session data - indeed, into all forms of activity data.

A session may encompass more than a single search; users, especially librarians, but possibly end-users as well, may stack queries for convenience and economy. For this reason it is important to distinguish between a search session and an individual search. In addition to containing a number of individual search statements, a session might also feature the changing of files and possibly, hosts, too. (Although the latter would be difficult to monitor, for it might mean matching different systems' logs.)

Number of searches conducted

For operational purposes an on-line search can be defined as that activity associated with a single query. It might, of course, involve a multitude of individual transactions - the use of various commands, displays of indexes and documents, and alternative search statements. In practice it is not always easy to distinguish between a search session and an individual search, because a user might use a variety of very different search approaches to get at what they are looking for. In the context of this research, it was judged that, where the terms looked very different, then this was taken to be a sign of different searches, and especially so if the exchange of terms was accompanied by a change in file. The problems encountered in session identification (i.e. when one starts and finishes) are also problems here, too. Throughout this investigation the

individual search has been taken to be the basic unit by which the use of other search characteristics, like file use, are measured.

Connect time

Connect time is probably the most obvious manifestation of use - and it is one that certainly preoccupies many researchers, but determining what can be read into the amount of time people spend on-line is a real problem. It is probably all too easy to fall in the trap and see speed as a component of skill - the faster the search, the more accomplished the searcher. On this basis end users could be expected to be slow searchers - and this is what some researchers have found (Lancaster *et al*, 1994). Plainly, this is not always the case for users, like journalists might be constrained in the amount of time they have for searching; also, someone with a limited grasp of the system is unlikely to search for very long without getting frustrated or humiliated and wanting to come out of it. The one - and only - thing that can be said for certain, is that in most cases quick searches are going to be cheaper. Hence the association with speed and prowess. There are two time-related measures of on-line activity: (1) the total amount of time spent on-line over a given time period; (2) the duration of individual sessions or searches. While total connect time is a use indicator, the duration of searches tells us something more, and exactly what else is open to debate: the aforementioned level of expertise, type of search conducted (e.g. comprehensive) or searching style (e.g. browsing)?

Much of the connect time data for POLIS were of limited value because many of the in-house POLIS terminals were left on all day long, with little attempt being made to log off from the system. This was largely because of convenience, and the frequency and urgency with which searches had to be conducted. With dial-up access and commercial hosts, connect time is a far more meaningful measure. Fortunately, this study featured two commercial dial-up services - TEXTLINE and FT PROFILE, and even in the case of the third on-line system featured - POLIS, an in-house system, there were people who dialled into the system.

Number of records displayed

The number of records displayed as part of a search provides relatively hard evidence as to the use made of an information system. In the absence of posting counts - and these were not supplied with either set of logs - they may also be used to measure the amount of data retrieved by end-users in relation to an individual search. Patently, the number of records displayed, is not necessarily the same thing as the number of documents found by a search, although it could be. However, it is always likely that

users retrieved more documents (postings) but chose to display only a proportion. Especially in the cases where large numbers of documents were found. Certainly you cannot display more documents than you found. In the particular case of PROFILE it was not always possible to establish the actual number of documents displayed. This was because users did not always indicate in their display statements how many records they wished to view, especially so in the cases where users specified **all** in their display statement. Therefore PROFILE records displayed figures must be regarded as minimum figures. In the case of in-house POLIS, where the monitoring software could not distinguish between an active search session and a period when the system was on-line, but at rest - thus placing a question mark over the connect time data, the records displayed figure gave the best evidence of system use. The display of documents is a conscious, demonstrable and quantifiable act.

Whether the number of items displayed provides, additionally, a measure of satisfaction or success of a search - as is claimed by some researchers (Barbuto and Cellavos, 1991) - is not quite so certain. Because the display command is used to screen for relevancy, items displayed might be a better indicator (than postings) of how many items a user wants in regard to a query. Furthermore, successful searching - especially in the case of full-text systems - requires that documents be retrieved and then displayed: though a zero posting might indicate that no material was available on a topic - possibly a positive result in the case of a journalist checking to see whether anybody had covered the topic that they were interested in writing a story on. Also, if a search finds nothing, then nothing can be displayed. Even if zero hits are taken to be an indicator of failure at the terminal then it is not necessarily always the user's fault. Failure may also be attributed to: inadequacies in database coverage, failure to add new material to the system in time, inadequate training or poor system design.

There is a strong presumption in the published literature that more documents mean higher degrees of success. But does it? Take the case of one user who typically displays, say, 50 records and another who averages 10. Is the first a better searcher because they find more, or are they in fact the poorer searcher because their broad and ill-defined searching produces more documents and noise. Perhaps, they were just doing different kinds of searches - the fact that one was doing an idea-generation one and the other a fact-finding one could account for difference. Dividing the number of records displayed by connect time provides a useful measure of productivity or search efficiency.

As well as success, large displays of documents - something associated with end-user searching in professional folklore (allegations made by *Guardian* and House of Commons librarians), are held to be a sign of a number of other things - searching style (browsing), retrospective searching (with the reverse chronological displays most systems default to the display of lots of documents inevitably means going back further in time), and the need for thoroughness or comprehensives in searching. The last association is the one that has captivated most researchers. Barbuto and Cellavos (1991), for instance, defined a comprehensive search as one that retrieved more than 50 records - a definition very much grounded in academia, and why 50 you might ask and not 30? Indeed, the validity of this number was called into doubt when the author observed that the number of records needed by users saying that they required a comprehensive search was scrutinised - a third of them were looking for 1-10 records. Comprehensiveness is of course a relative term and the attempt to measure it by document numbers alone can result in some not so convincing statements.

Number of screens viewed

A variation on the number of records displayed is the number of screens the user views as a result of a search - something which could be identified from the PROFILE logs. While an interesting measure in its own right, it is another rough and ready activity measure. You cannot equate screens viewed with the number of records displayed, because the number of records displayed per screen is variable, being dependent on a number of factors: (1) the format chosen - the briefer the format the more records contained on a screen; (2) the length of the record - the longer the record the more screens needed to view it; (3) and, because it was a Carriage Return that signals a new screen, if users choose to override the natural break (of 18 lines in the case of PROFILE), with a **Nobreak** command then there is no knowing how many screens they viewed.

Number of on-line interactions

In the analysis of on-line use, a distinction has already been made between a search session and a search. It is also possible to make a distinction between a search and the component parts of the search: individual interactions or transactions. Thus a single search session might involve a large number of individual interactions with the system: terms may be input, added to or changed or, perhaps, the index is examined to check on term availability. The attraction of using the number of operations as a measure of activity is that it provides a measure of busyness when on-line. As Peters *et al* (1993: 45) point out *two search sessions may last the same amount of real time, but one may feature a lot more commands than the other ... two searches may contain*

the same number of commands yet one could last substantially longer than the other ... .The precise nature of the search operations used is something taken-up later in the examination of command usage.

Measures of satisfaction

This is probably the most controversial on-line measure. What does it mean when a searcher says that they are satisfied with a search - presumably that they found what they were looking for or something like it. Inevitably satisfaction is bound up with success. Use itself is, of course, some measure of satisfaction. A system that failed do deliver satisfaction would soon fall out of favour. For some users satisfaction might be met in the one document that provided the answer, and yet for others - typically academics - satisfaction comes with completeness, resulting in possibly dozens of documents. The trouble is that satisfaction is a relative and very slippery term. Low expectations and skills mean that satisfaction may be achieved fairly easily. For end-users the ability to find relevant information with simple search techniques, with little expenditure of time, perhaps in the comfort of their own offices is always likely to inflate satisfaction scores. Novelty must play a part too. On the other hand - and this largely concerns professional or experienced searchers - the better you are at searching perhaps the more conscious you are about your own frailties - and hence not wholly satisfied with your searching.

Whether it was a *good* search, in that it uncovered everything that there was, or the best document that there was, and was conducted speedily and efficiently in terms of appropriate term and command use, is a different matter (and addressed below). With commercial and dial-up systems cost must surely enter the equation, too. If end-users are aware at all of what constitutes a good search - and they are hardly ever put in a position of assessing or knowing themselves - they probably obtain that awareness from examining what intermediaries have done for them. But whether that makes them less satisfied with their own searching is open to doubt: if for instance the end-user found 5 relevant records when the library was closed or busy or couldn't respond in time, would it really matter that the librarian would have found 20 and maybe better ones? The answer is probably no, in the case of the politician or journalist.

For many of the reasons described above Lancaster *et al* (1994) believe that user satisfaction is a wholly unreliable measure of on-line success. They claim that users are all too readily impressed by computerised information retrieval and are all too easily lulled into *a false confidence syndrome.* They point to the many surveys that

say that users are satisfied and yet by all the technical (system) measures of performance they search poorly. As far as Lancaster *et al* (1994*)* are concerned the only valid measure of success is one based upon recall and precision scores. However, they must have been thinking solely of bibliographic systems, for it would be virtually impossible to undertake such an analysis on a full-text system like PROFILE, where the search possibilities are endless and the target document population unknown. They were not thinking of journalists either, for they do not always set out with a target in mind. And there is certainly no question of them putting the user first in their deliberations.

Apart from arguments about its worth, satisfaction also poses problems when it comes to measuring it - and for this reason it is said that this is not the domain of logs, but the domain of surveys. Of course, computer logs give no direct evidence of how successful or satisfied users were with their searches, but the logs do in fact provide some clues. Firstly, there is the number of records displayed mentioned earlier. Something positive can be read into the fact that records were actually displayed as a result of a search. However, this does not merit as a strong measure of satisfaction, for allowances have to be made for the fact that negative searches can be positive, and that a display of records does not necessarily mean that useful records were found. Nevertheless some researchers are beguiled with the simplicity of the method - thus Meister and Sullivan (1967) defined a successful search as one in which at least one document is displayed.

Secondly, in respect to full-text systems, like FT PROFILE, there is also the evidence as to whether a search ended with a full text display. The bibliographic analogy being going to the shelves and retrieving the document. This is probably the stronger measure of the two, because a display of records in their headline form might only provide confirmation that the documents were not relevant, whereas a full-text display usually follows a **context** or **headline** relevance check. Presumably then, the more full-text displays the higher the yield of relevant documents, and the higher the degree of satisfaction. This indicator is not completely free from reproach, for it is possible for users - say, those looking for a fact - to obtain the data they wanted from a paragraph display only.

Measures of expertise

On-line expertise or skill can be demonstrated in a variety of ways. The traditional way of determining this has been to examine the range of commands being utilised.

The underlying assumption being that the more commands used the more expert the searcher. This is easily determined, as too is the number of input errors made (when entering commands and terms) - another possible indicator. The other possible signs are not so easily assessed, they are: the appropriateness of the commands used; the structure of the search query (number and quality of terms, construction of statement); the appropriateness of file selection and willingness to change files; the development of the search strategy as a result of interaction with the data - the number of search steps might prove an indication of this. But as has already been pointed out, on-line expertise and satisfaction with the product of a search does not necessarily go hand-in-hand. It is well established that end-users with minimal skills still achieve high rates of satisfaction (Dalrymple, 1989).

Commands used.

The supposition is often made that the wider the range of commands the better the searcher, the more effective the search, and the better the search result. However, Standera (1975) believes that there is nothing wrong with the use of a small set of commands, and that it could be quite cost-effective. Perhaps, too, commands are not so much under-utilised but just not appropriate, just extra-baggage as far as the end-user is concerned. It must also be remembered that when it comes to the number of commands utilised this must be related to the number of commands available on the system. To say that one command attracted 50% of all use in a four command system is somewhat different from saying the same thing about a ten or twenty command system. The two principal systems featured in this study - POLIS and PROFILE - are poles apart when it comes to command availability: POLIS has probably as many as any system and PROFILE has less than any other system. Some commands are undoubtedly more difficult to carry out than others - and the use of these advanced commands is also scrutinised by researchers to detect signs of end-user weakness. Boolean searches, truncation and limiters are typically mentioned in this context. But what is expertise to be measured against, what is to be the benchmark? Well, in most cases it is against professional information workers. And this benchmark is used throughout this study. Though it has been shown that even practised and skilled information professionals show little agreement as to search strategies and terms chosen (Saracevic *et al*; 1988; Fidel, 1985).

Errors made

Using wrong or unsuitable commands, entering them in an illogical order, or typing them in badly, can all signal weaknesses in the searcher. However the judgement as to whether the commands were inappropriate in the circumstances is difficult, for there

are so many routes to on-line searching success, especially in the case of a user who is not exactly certain of what they are looking for and who is searching a system, like FT PROFILE, which has very little structure. Researchers are on stronger ground when looking for spelling or input errors. The PROFILE logs highlighted the problems of spelling people's names. For most researchers though evidence for search weaknesses is to be found in the retrieval of nil sets.

File selection

Single, blinkered and inappropriate source selection is often thought to be a hallmark of end-user searching (Oppenheim, 1992). Equally, the judicious use of files and file switching is held to be the hallmark of the expert searcher. From logs, it is easy to spot file changes and monitor the number of files selected, but it is not so easy to determine whether the files selected were appropriate. Multi-file groups complicate matters somewhat, for what can you read into a search of a tailor-made file group of some 28 titles, like PROFILE's UKNEWS. Indeed, there is a siren-like quality about these groups for end-users and professionals alike, and it could be argued that selecting such groups is a little bit like falling of your chair (too easy) - and thus represents a lower form of searching. In systems like POLIS and PROFILE via ATEX, were the system rests in a particular file, this must influence source selection too. File use can be profitably examined in two ways: by counting either the number of searches conducted in them or by counting the amount of time spent in each.

Time of searching

Logs provide a good deal of detail in this regard, enabling time, day, and month of searching to be determined. Time features strongly in many log analyses. We have already mentioned it use as a means by which use and searching efficiency can be measured, and it can tell us something about the pattern and spread of searching. Of special interest here is the assertion that end-users leave searching to the last minute: just before the Chamber sits in the case of The Commons, or in the case of journalists just before the paper is sent for printing, in the early evening. Dial-up users are also of interest, for they are in a position to extend the information seeking day - do they avail themselves of this opportunity?

Nature and types of searching

The dynamics of the searching process can be examined. By looking at the sequence

and mix of commands used it is sometimes possible to chart how a search progresses, unfolds, and changes. In connection with this the following search characteristics have been studied: phases in searches - exploratory, backtracking and error stages are some of the ones that have been mentioned in the literature; the broadening or narrowing of searches; the changes in 'search state' - the move from a subject to non-subject searching for instance; and the interplay between controlled language and natural language searching. Neither of the logs examined in this study provided enough evidence to assess these characteristics, or where they did it was too difficult to confidently interpret the data. And, anyway, such analyses could only be done on a very small-scale. But it was possible to investigate: the access points used, whether a default search was used; and how many stages the search passed through. What was especially being looked for was: what were the popular search approaches adopted by end-users, whether the full range of fields were utilised, whether search statements were completed in a single line, were searches broad or narrow; and whether they were simple or complex in construction.

3.3 Interviews and questionnaires

These two staple social science research methods are best employed to measure attitudes and opinions towards on-line searching, and to assess perceptions of success. Both methods were employed to the full in this study to canvass the views and opinions of end-users, potential end-users and librarians (see Table 3.1).

Interviews

No research method is entirely free from problems, but interviews must surely come close. Interviews can probe for quantitative data in regard to system use, help identify problem areas, and provide an understanding of the context in which searching is undertaken (seeing people in their work environment is very important to the understanding of their information needs). Furthermore, interviews have a habit of throwing up the unexpected, things that were not asked about (but with hindsight needed): with a good interview, the interviewee determines the scope and nature of the questioning domain as much as the interviewer. A rich vein of data can be tapped from interviews, especially when conducted shortly after a search or if conducted with regular users. And of course there is no real substitute for the method when it comes to studying non-use and non-use was studied in the context of both journalists and politicians.

On-line searching is not a self-contained event, but a particular form of information seeking triggered off by a need for information. Therefore to understand it, and all that goes on as part of the search, the event must be viewed in the broader context of information need and information seeking. Needs characteristics, such as: the requirement for very current information and speed of delivery, the size of the information appetite, whether highly processed information is wanted; and general constraints on information seeking, like lack of time to search and digest the information, are all highly significant in understanding on-line searching behaviour - and needs data can only really be satisfactorily obtained by interview. Lancaster *et al* (1994) really miss the point when they bemoan the lack of genuinely evaluative studies in the field - and attribute this to insufficient attention to recall and precision, when in fact what many studies really lack is an information needs' dimension.

Both the case studies of on-line searching behaviour presented here are grounded in an analysis/understanding of information needs and as a result the interview is a prominent feature of both case studies. Journalists and politicians make very good interviewees, because they have no qualms in telling information professionals and researchers exactly what they think. They are generally full of confidence and self-esteem and do not feel that they have to conform to the expectations of the interviewer, and they are also exceedingly good at expressing themselves. In the case of both journalists and MPs questioning was open-ended and free flowing: a stylised, specific question approach would have not worked for it might have put non-users on the defensive and would have tested the interviewees patience (see the respective interview schedules - Appendices 1 and 6). In the case of *The Guardian* it was possible to undertake repeat interviews with the same people, to chart changes in attitude and use. The chief problem though was in pinning users down to a time to interview them - both groups being incredibly busy. Because of this interviews with politicians and journalists were often undertaken at lunch, over the phone, or as part of a lecture. Altogether 10 MPs and 18 journalists were interviewed for an average of one hour each. In addition 15 Research Assistants were given interviews - these interviews were based upon the self-administered questionnaire that was sent to their colleagues. Library staff from *The Guardian* (8) and House Of Commons (22) were also interviewed (see interview schedules in the appendices). Given the regular contact the researcher had with both sets of librarians interviews were conducted somewhat more informally.

Questionnaires

Inevitably, if large numbers of users are to be surveyed and resources limited, then

time dictates that questionnaires have to be resorted to. But in some ways they are unsuitable for the task in hand, because: (1) when you are asking about databases and searching commands it is very difficult to formulate questions that are completely free from jargon (asking about truncation, for instance); (2) it is very difficult to be certain that users understand the line of questioning and ambiguity frequently occurs; (3) response rates are notoriously low, especially when busy practitioners are concerned (bulging in-trays tend to compete for attention and bury questionnaires); (4) attempts to obtain statistical data can descend into vague and subjective categorisation - satisfactory, very satisfactory etc. However, a well piloted questionnaire, sent to users adept at annotating and editing, which has a good internal chasing system (as was the case in *The Guardian* survey) and one in which the respondent feels that there will be some immediate and significant benefit, can produce rich dividends.

One way of getting away from the rather broad or generalised opinions that surveys typically provide, is to conduct the survey closer to the point of searching (Hancock-Beaulieu, 1990B). That way more detailed and reliable data on individual searches can be obtained. Questionnaires can be distributed before and after, and changes in search strategies or requirements as a result of going-on-line can be detected. Attempts have also been made to remove the influence of the interviewer by introducing machine-generated questions as the search progresses. The method is still intrusive and is of course potentially disruptive. However, is seems to have had some success with City University researchers, who have conducted experiments with an *interception front-end* called OLIVE (Ibid.: 53). For busy politicians and journalists such methods would have proved too disruptive. At The House of Commons they would not even agree to get staff to log off after every search - to improve accuracy of the logs - because they felt that would slow their response down unacceptably.

The questionnaires for both studies tackled much the same ground, but they did differ in some important aspects. *The Guardian* journalists' questionnaire asked more detailed questions about the use of search commands and file selection. This was because it was by no means certain that logs would become available - they thus had to double-up as a record of search behaviour as well as a method of determining attitudes. At the House of Commons the log and questionnaire studies were being conducted in tandem so there was no need for such detailed questioning. Of course users were being questioned about different systems so naturally there would be differences in questioning. What did restrict the room for questioning in the case of the The House of Commons was the need to question the users about two systems - POLIS and TEXTLINE. Ever present in the researcher's mind was the time it would

take very busy politicians to fill in the questionnaire - and whether they would spare this time. The phrasing of questions differed, too. This was partly because they were tailored to the precise circumstances that prevailed in each institution at the time of distribution - at *The Guardian*, for instance, journalists were experiencing access problems with only two passwords to share amongst them and delegation was an issue that particularly concerned librarians at the time.

Although very similar numbers of questionnaires were distributed to the two groups of end user, the numbers answering the questionnaires and the response rates differed also. At The Commons 60 Research Assistants (52% of those asked) responded and at *The Guardian* the numbers and response rates were even higher - respectively, 87 and 73%. A higher proportion of journalists were sent questionnaires because of the uncertainty of obtaining the logs. Journalists' response rates were higher because of two reasons: (1) the *Guardian* library staff were on the spot, and they wanted the data as much as the researcher so they actively chased the journalists; (2) at The Commons the Research Assistant population was a shifting one - and as a result it was difficult to establish an accurate sampling frame.

3.4 Observation

The need to witness on-line use in the broadest information seeking setting and to see the live interaction between user and information system - in the case of the librarians between user, intermediary and system, meant really only one thing: that users be observed in a work setting. Because of the different circumstances that prevailed at the two host institutions two different approaches were taken. At *The Guardian*, where security and confidentiality concerns were not so great and where journalists worked in an open plan environment, it was possible to observe individual end-users - three were informally observed for a day. As well as any contact they had with on-line system, their general information seeking behaviour was also observed to provide a context for the searching. In addition the entire City desk - six people at the time, was observed (in exactly the same manner) for a day. This was done because they were the most active users and, conveniently, they all sat in the same room. Observation in these circumstances was bound to be productive.

At The House of Commons, where access to individuals' offices was not so easy and where the decentralised nature of workplace made observation somewhat more difficult, arrangements were made to observe politicians and Library staff at work in two locations: in The Oriel Room, the main point of contact between librarians and

MPs. Observation there spanned two half-days and covered three library staff and 14 politicians; and in the Branch Library, the place where most of the on-line searching went on - the period of observation being a day - the on-line searches of 12 Researchers being witnessed. In the case of the Oriel Room, observation was followed up by questioning of staff to obtain clarification of the actions witnessed.

3.5 Other methods

Talk-aloud techniques

Lancaster *et al.* (1994) maintained that the real interest for researchers of on-line searching behaviour lay in why certain approaches were adopted and others rejected, and as a result, were more inclined to getting users to verbalise their thoughts at the terminal, though they readily admitted that it could be clumsy in execution and could interfere with the search process (and failed to use themselves). Also a very high degree of co-operation is required from the users. In many ways the weaknesses of the verbal method are very much the strengths of transactional log analysis. Sometimes researchers use other labels to describe the method as protocol analysis or verbal data methods. Dalrymple (1987), Sullivan (1986) and Hancock (1987) have all used the technique to seemingly good effect with groups of academics, but there is probably limited mileage for the technique outside of the laboratory that is the university library. A brief attempt to try the technique at *The Guardian* soon alerted the researcher to the fact that few practitioners would be willing to expose themselves in such a way; that they also searched at unpredictable times and in a rush, which would make monitoring difficult and, possibly, would not want to verbalise their search because of confidentiality concerns - journalists work in a competitive and open-plan environment. One can see why researchers are fond of the method because it provides unique and invaluable data - on the formulation and changes in search strategy - that other methods can only really guess at.

Recall/Precision tests

Such tests have a long-history in measuring the success of searches and are largely associated with the evaluation of systems. But the presumptions behind them are rather shaky, particularly when considered in the practitioner environment. Because of the inverse relationship between the two, there is often a need to plump for one or the other: recall has proved the librarian-favoured measure. But this is a measure of success more likely to be accepted by a researching academic, determined on

completeness (pinning down all the potentially relevant literature). Fraught and busy practitioners, with deadlines of hours, if not minutes, using the facility as a supplement to their other methods of information acquisition are unlikely to see it that way. Given the aforementioned, it might be expected that practitioner end-users would score more highly in precision terms and look more favourably at precision as a measure of success, but the way that many journalists and politicians generally seek information - by browsing around - means precision scores are not going to be all that significant either. Recall/precision methods assume that you know what you are looking for in the first place and that others can identify and establish exactly what you should be finding, but the serendipitous route to end-user success and the difficulties of establishing target populations in full-text natural language databases means there are going to be problems in the computation of the scores. Lancaster *et al* (1994) overcame some of these objections by allowing the users to determine the relevance of both their searches and those of the librarians, but this is much easier to do in the context of a bibliographic database, where populations of relevant documents are much more easily assembled, and in the light of the much more focused needs of academics.

CHAPTER 4
POLITICIANS AS END-USERS
CASE STUDY: *THE HOUSE OF COMMONS*

At the House of Commons they have been searching on-line databases since 1977, when they took out a subscription to BLAISE, although in-house end-user facilities were only introduced in 1986 and dial-up ones in 1987. During the period of study two on-line systems were made available for politicians and their staff to search themselves: the Parliamentary On-line Information System (POLIS), which was then essentially an internal library catalogue; and TEXTLINE, a remote commercial newspaper cuttings-style host. The period of research encompasses the years 1988-1990, although a study by The House of Commons Information Committee (1994) updates some of the data provided here. Thus this case study differs from the second one - the journalist one - in at least two respects: (1) it covers two on-line systems; and, (2) it spans a shorter time period. Indeed, there is a third difference - and this compensates in part for the more limited period of the study: the numbers of potential on-line users under investigation was much larger - at The Commons there were more than 600 MPs, 600 MPs' Research Assistants, and a Library staff of around 150 to compare them with.

The fact that two very different information systems were being searched by potentially large numbers of politicians and Library staff meant that there were greater opportunities for study and comparison. Thus it was also possible to study the differences between: in-house and dial-up users of the same system (POLIS could be searched by end-users using in-house or remote facilities); individual Library sections (was it possible, for instance, to generalise about intermediary searching?); the use of a bibliographic system (POLIS), largely used for the identification and location of documents and a text-based system (TEXTLINE), used largely to provide information and answers to questions. Also, with both TEXTLINE and POLIS searchable in both menu-driven and command-driven forms, another interesting variable was introduced into the examination of end-user searching at The House of Commons.

4.1 Research Methods
A number of methods were employed to monitor and assess on-line searching at The House of Commons. This was necessary because no single method could be trusted to yield a complete or wholly reliable record of end-use. Central to the study were

computerised transaction logs obtained for both POLIS and TEXTLINE users. The POLIS logs, produced by the BASIS software were very comprehensive and provided the following data for each of the system's password holders:

(1) twelve months (Sept. 1989 - Aug. 1990) worth of search session details: yielding information on the number of search sessions conducted; the duration of sessions; the number of records displayed; the time when searches were conducted; and the files searched.

(2) twelve days worth of information on the search commands utilised, fields searched and terms used. The data was available for the busiest day of the last ten working days for each month in the year.

TEXTLINE logs were not as complete or as detailed as those for POLIS though they still furnished information on twelve months of end-user searching (Oct. 1989 - Sept. 1990). For TEXTLINE the following data was available: number and duration of search sessions; date when searches were conducted and file choice.

Because the logs for both systems related to terminals rather than individuals, and because each terminal had a variable number of users, comparisons between sets of users have to be treated with some caution. It was not always possible to tie searches down to individuals as was the case with *The Guardian* study that follows. In fact, one condition of access to The Commons was that was not done.

In the case of the politicians these logs were supplemented, and enhanced, by a questionnaire and interview survey and observation. The function of the survey was to: (1) obtain an understanding of politicians' information needs and place on-line searching behaviour in this context; (2) obtain background data on the user group; (3) ascertain end-users' attitudes and opinions towards on-line searching - about which the logs could say little; (4) cross-check and enhance the data obtained from the transactional logs. For Library staff members - a smaller and more accessible group - interviews alone were undertaken to supplement the transactional logs.

Through preliminary discussions with Library staff it had been established early on that Members' Research Assistants were the principal end-users at The Commons: Members, with the notable exception of Enoch Powell MP, readily delegated all their formal information seeking tasks. In some ways this discovery was very fortuitous, because it would have been far harder to obtain the co-operation of MPs than their researchers - MPs are typically too busy to co-operate in research studies, whereas their researchers, a largely unsung group, were generally only too happy to co-

operate. Nevertheless, ten MPs were interviewed to confirm this fact, and also to obtain information on their information needs and their relationship with their political staff. At the behest of The Commons' librarian these interviews were held away from the House - mostly at The University of North London. The format of the questions were open ended, focusing first on the routines of the job and then drawing out the information need implications that arose from the descriptions (*see* Appendix 1 for the interview schedule). In two cases interviews had to be conducted over the phone because it proved impossible to meet up with MPs.

At the time of the study 600 people were classified as Research Assistants by The Commons authorities - too many to study individually. Fortunately, many were political researchers in name only, undertaking largely secretarial activities. Even with these weeded out (The Library proved invaluable assistance here), it was still necessary to sample the population of researchers. A 20% sample was seen as reasonable and achievable goal. The sample was drawn up using a number of methods to overcome the lack of an up-to-date, accurate sampling frame (there was a large turnover of temporary Research Assistants) and to ensure a representative sample was obtained. It was especially important to ensure that all the various types of Assistant were represented: the permanent ones, who largely worked for Frontbench spokesmen and had (limited) access to the Member's Library in the Palace of Westminster; and the temporary ones, who largely worked for Backbenchers and had only access to a (geographically) remote Annexe of the Member's Library - the Branch Library.

The sample was drawn up as follows:
- Firstly, Research Assistants who signed on for the use of POLIS and TEXTLINE at the Branch Library between March and June 1989 were identified (a log book was kept by the side of both terminals for this purpose). Within this group, a smaller sample was randomly selected for an interview: the first few interviews were used to pilot the questionnaire. The rest were sent questionnaires. This exercise would have picked up temporary Research Assistants, US students on placement among them.
- Secondly, a group of permanent Research Assistants (of which there was a record), who had access to the Members Library and who had not used the Branch Library terminals during the survey period, were sent questionnaires.
- Thirdly, dial-in POLIS users, who had not been picked up by sampling stages 1 and 2, were also sent questionnaires. These people, most probably Research Assistants - but just possibly Members, were working from their constituencies or from personal computers in their offices in the Palace of Westminster.

- Fourthly, Research Assistants using the Branch Library, though not necessarily using POLIS or TEXTLINE at the time, were given questionnaires during a spot survey, which lasted one day.

A sample copy of the questionnaire used can be found in Appendix 2. The questionnaire was also used as the interview schedule, though interviewees were not confined to its boundaries, and if the interviewee raised related issues these were noted.

In this way 115 research assistants were approached and responses were obtained from 60 of them, giving a response rate of 52%. Of the 60, 15 were interviewed and the rest completed questionnaires. Because of the manner in which the sample was compiled the sample population was biased towards users of on-line databases. Whilst this sample was obviously slanted, in the context of this study this is probably acceptable, as the prime focus of the study was the manner in which on-line databases were being used.

In the case of Library staff, one hour interviews were held with approximately one third of the staff (22), selecting only on-line searching staff, who had a direct responsibility for answering Members and their staffs' queries (Appendix 3 for a sample schedule). Staff were also chosen to obtain a representation across the various Library departments and sections.

There were only limited opportunities for observation at The Commons as security and confidentiality were always very pressing concerns with the Authorities. But where they presented themselves the opportunity to witness on-line searching in the round was always taken. As a result, an opportunity to conduct a two day observation of Library staff searching POLIS on behalf of MPs in The Oriel room - MPs' major point of contact with the Library did arise. A brief schedule helped direct the researcher's attention to matters of interest (Appendix 4). Fourteen MPs and three Library staff were observed during the two days.

4.2 End-Users at The House of Commons

At The Commons two groups of political end-users had to be considered: MPs Research Assistants, the main focus of this study, and the MPs themselves, who turned out to be largely bystanders in the on-line process.

Research Assistants

There were two main categories of Assistant: permanent and temporary. There were approximately 200 permanent staff, who had a large '9B' stamped on their white photo identity card, which entitled them to use the Members' Library of the House of Commons at restricted times (during two mornings) and 400 temporary ones, who had a 9D pass and who could only use the Branch Library.

It is hard to generalise about Research Assistants (as it is about MPs themselves) and even within each group, there were variations, but, broadly speaking, there were those who worked for Shadow Cabinet Ministers or Opposition frontbenchers, and those who worked for Backbench Members. Frontbench Research Assistants were mainly permanent staff and were usually highly experienced researchers, with an academic background and some subject expertise in the area in which their employers specialised. They worked closely with their employers and were often engaged in political research and writing. Backbench researchers could be based in the House or in the constituency. Sometimes they worked in both places. The former were mainly concerned with drafting Parliamentary Questions, preparing briefing papers and doing some background research while the latter (usually political appointees) dealt with constituency casework, and even held surgeries.

Amongst backbench Assistants, there were a number of identifiable groups. One group was employed to work on specialised projects, for example, the Channel Tunnel rail link. They were usually employed on short-term contracts by professional bodies and pressure groups, and did not use the House of Commons Library very often as they had other sources on which they could rely. Another group of researchers were students, temporarily working in The House as interns. Many of them were American undergraduates, who worked for a British Member of Parliament for 6 to 7 weeks as part of their degree course. The students were usually engaged in drafting letters and producing information briefs. Because of the pressure on space at The Commons and the considerable demands they made on staff time, their numbers were limited to 50 at any one time. Typically, this group knew little about the British Parliamentary process - and it was alleged to show in their on-line searching. British university students also worked for Members of Parliament, either in the summer or on placements as part of their degree courses.

The turnover rate for Research Assistants was high - ten to twelve months was the average tenure, which meant that few had the time to become proficient Library or on-line searchers, something which has to be considered in the following analysis.

Complicating matters, many Research Assistants worked part-time and a number also worked for more than one Member. Most of them were committed to their work but tended to be young and inexperienced and regarded their work experience as useful and a good way of beginning a political career.

The type of work that a Research Assistant did depended largely on his/her Member's needs - they were the tool of the Member for whom they worked. Some Members only wanted someone to help them deal with correspondence and check a few items in the Library, while others would delegate a great deal of the information and research work to their Assistants. These were some of the tasks that Research Assistants were expected to do: correspondence and filing; typing and word processing; preparing press releases and Parliamentary Questions; writing briefs and reports; research and presentation; and meeting constituents. It was estimated that about 60% of their time was spent on information and research work. Information and research work included writing reports, speeches, press releases and research briefs, and obtaining information from newspapers, journals, Bills, etc.

Research Assistants had some grounds for complaint when it came to Library matters, for their information-seeking activities were prescribed and curtailed (*parcelled out and restricted* according to one Assistant). The Library had, if not a policy, a strong preference for dealing directly with Members. According to another Assistant: *The House of Commons system makes it`very hard to get assistance because it operates on a priority basis, with Members' requests being dealt with first.* This factor is very important in understanding the way they use databases for they provided them with an opportunity to by-pass the obstacles placed in their way by a very bureaucratic system.

Of course, not all Research Assistants searched databases themselves, though they did constitute a large and very active on-line group, the biggest in The House of Commons. Of the 60 Research Assistants surveyed, 57 used at least one of the two on-line systems placed at their disposal. From the comments books in the Branch Library, which listed POLIS and TEXTLINE users, it was calculated that about 100 (17%) of the Research Assistants were engaged in on-line searching. Additionally, another twenty two had dial-up facilities in their offices or constituencies. The membership of the two groups did overlap, but that probably still left around 120 (20%). Assistants who were actually engaged in on-line searching at the time. By any count then, they represented a major force of end-users. Indeed, at the time there were as many end-users at the House of Commons as at Rupert Murdoch's News

International, that bastion of end-user searching in the UK.

Members of Parliament
Although MPs themselves hardly featured at all in this end-user study - after talking to Members, Library staff and on-line hosts (PROFILE and TEXTLINE), it was established that the population of (sporadic) on-line using MPs was thought (nobody knew for certain) to number at the very most less than half a dozen. None the less, it was still necessary to consider this group briefly to establish the levels of on-line awareness and the reasons for non-use.

From the interviews conducted with MPs, their Research Assistants and Library staff, it was clear that Members showed very little interest in the on-line searching process. They appeared to know little beyond the fact that Library staff and their Assistants were using computers to retrieve data - and even then they jokingly referred to the computers as *that machine, black box, screen*, etc. In the opinion of a Commons' Librarian: *They wouldn't know one database from another; They wouldn't know the name of the database being searched, nor could they distinguish PROFILE from POLIS - indeed, because it is all on the same screen, they probably would think it's all the same*. To be honest Library staff did not really help here and must take some of the blame because they were reluctant to search the databases in the presence of Members, in case MPs intervened in the searching process or, more likely, saw errors being made. A few Library staff were of the opinion that computerisation had meant that Members were even less likely to help themselves to the Library's contents: *the more the Library staff use and employ technology, the further away from it Members will get and are, in consequence, far less likely to use it (or question its use)*. An interesting rebuttal of 'the more you see computers being used the more you will use them yourselves' argument. And one of the MPs interviewed did say that he regretted *the passing of the manual information systems*.

In general, Library staff thought that there was little chance of Members ever becoming end-users. They put forward a number of reasons why they thought this - most of which referred to the general willingness of Members to delegate all formal information seeking.
1. They have no time to search themselves - they are too busy, with too many demands on their time and much to much to get through. Members would not have the time to learn the techniques: *for Members, their lives are a constant struggle to keep up with things; they simply haven't the time to search POLIS - menus or no menus*. (A number of the Member's interviewed read out their

week's engagements - and it can be confirmed that there was little spare space to do anything.)

2. They would find the subject approach very difficult to handle: *it's a complex system*. This is, of course, an allusion to the problems of searching a controlled language, bibliographic system. In the survey of Research Assistants this difficulty was acknowledged by many of them (section 4.3.4), but it did not stop them from searching POLIS.

3. Members have no wish to learn - and remember - the mechanics of information retrieval. A Library staff member summed up the situation in this way: *POLIS is not really appropriate for direct use by Members. They wouldn't have the time to use it regularly and hence become proficient in its use. They would come back after the long summer recess having forgotten all about it*. But their Assistants didn't forget.

4. There was always someone there to do the searching for them - and at very short notice. Many Members held senior positions in public and private organisations, and were simply used to delegating. Perhaps, the most telling point of them all.

5. Space precluded it anyway. There was not much room in the Library or in Members' offices for on-line terminals - *if they had space for a desk, they'd use it for a desk*.

6. In the particular case of POLIS, there would not be a lot of incentive to master what is essentially a bibliographic system, which just provides sign-posts to documents. If it was full-text, that might be a different matter. *As it was, someone has still got to fetch the document*. But then none appreciated the (largely) full-text TEXTLINE either or as we shall see later on FT PROFILE (page 158).

7. For reasons of self-image or status, they do not see information retrieval being one of their responsibilities. This explanation reminding the researcher of the reason given to him by a senior businessperson for not searching databases themselves - *I know how to make coffee but I don't*.

8. Members were not computer (or library) literate, they feel very much happier obtaining information orally. *Members do not realise the printouts they get are from POLIS, or what POLIS represents. They do however make a point of knowing what a member of the Library staff can do for them*. This came out very strongly in the interviews: when two Members were asked whether they searched databases, it was clear from their blank looks that they had never heard of the term or understood the concept. Another MP responding to the same question said *no I am not particularly good with computers*.

Among the 10 Members interviewed, none showed any real knowledge or familiarity with on-line systems or POLIS. With two exceptions, they were aware that Library staff did search databases on their behalf and that the Library had its own database, but only two could identify it by name. Most though were aware and interested in the fact that access to POLIS could be obtained from their offices: *It would be useful to go into Prestel or (what's that one at The House of Commons called?), ah POLIS, without bothering the Library.*

Although Members appeared little interested in on-line systems, they have contributed to an increase in Members' enquiries - computers have raised the information profile of the Library. One of the prevailing problems encountered by many libraries - The House of Commons Library amongst them - is that users are generally unaware of the resources the Library can tap into on their behalf. Users commonly have a very limited and traditional view of what the Library can offer. As a consequence, they do not use the Library when maybe they should. Computerisation might be changing this: Members had an almost child-like belief that the computer (the black box) could answer anything. A situation that actually led some Members to believe that the computer would provide an answer to the question and not just lead to a document that might help in answering the question. So perhaps, just the presence of POLIS terminals (although Members do not necessarily recognise them as such) has actually led to an increase in demands made upon the Library.

4.3 USE OF THE PARLIAMENTARY ON-LINE INFORMATION SYSTEM (POLIS)

4.3.1 POLIS: the on-line system

POLIS is the House of Commons Library's very own computer database. At the time of the study it was essentially a bibliographic database of name-and-subject-indexed references to current and recent Parliamentary information. At The Palace of Westminster 50 dumb terminals were linked to the system. These terminals (since replaced by PCs) had programmed function keys to speed up the searching process (i.e. you did not need to type **browse** to get in the thesaurus, instead you just pressed the labelled function key). Most of these terminals were located in the (numerous) offices of the Commons Library, though terminals could also be found elsewhere - in the Table Office for instance. The terminals situated in the Commons (Members) Library were for the use of Library staff members only. One terminal, introduced in 1987, was reserved for the use of Members' Research Assistants in the Branch

Library - there was no end-user facility for searching POLIS in the main (Members) library, though in fact this library was closer to most Members' offices. Anybody in the Palace of Westminster, who wished to use a POLIS terminal themselves had to leave The Palace, walk about half a mile and pass through two security checks. MPs and their staff could also dial into the system from remote locations, free of charge, though they had to foot the hardware and software costs themselves.

POLIS is principally a parliamentary database and much of the material covered is of parliamentary origin: Parliamentary Questions (PQs), Parliamentary proceedings, Parliamentary papers and legislation loom large in the database. These materials accounted for 90% of all the records on POLIS, with PQs making up the bulk of records (58%). At the time of the research POLIS held no records of journal articles or newspaper stories - though it does now. It was, therefore, not an index to the Library's considerable journal and newspaper resources. As of September 1990, POLIS contained 841049 document records organised in seven databases: CURRENT, PARL83, PARL79, EDM, OPQ, HOLLIS and PRESS. Essentially, parliamentary material overflowed onto three databases - CURRENT, PARL83 and PARL79; that is one database for each of the three Parliaments since POLIS' introduction. On each database was the material produced by the House of Commons in that particular Parliament. CURRENT database - the largest of all - then covered the sitting Parliament (1987 to 1991); PARL83 covered the previous Parliament (1983 to 1987); and PARL79 (1979 to 1983) covered the one before that. The EDM database covered just one form of material - Early Day Motions.

PRESS was a database of government press notices received in electronic form from the Central Office of Information. This file appeared too late to be picked up by the monitoring logs. HOLLIS, which was a House of Lords database, covered pamphlets and official publications from 1979 to April 1987. It was a dead file with Lords Library's current pamphlets and official publications being added to CURRENT. Finally, there was the OPQ (Oral Parliamentary Questions) database, which was excluded from consideration, because it was purely administrative in function used to check question authenticity and originality.

In retrieval terms POLIS is relatively sophisticated - and much more so than FT PROFILE (the host featured in the journalists' study). It has a complicated field structure, a large controlled language and an impressive armoury of search facilities, in all this it was the very antithesis of PROFILE. You can search POLIS for a topic, limit your approach to any one of twenty-three fields, specify that you are looking for

general or specific (case) material, narrow the search to any one of 38 publication forms (surely, something of a record?). If you are uncertain as to what words/terms to look under, help is at hand in the form of an alphabetic and systematic (subject) index - the latter boasting some 9000 terms. And, if you choose the wrong controlled language term, the system may automatically correct the error for you and find the correct equivalent by using a switching device. You may also save your search, to be later recalled in another database, or, at another time. When you are finally ready to print out the results of your search, POLIS can sort the output into an order of your choice. If you are a novice searcher or end-user, you may, search the system in its menu-driven form - and probably best advised to. It has to be said that POLIS in the raw is not a system for end-users or the faint hearted, it can be intimidating, as members of the Library's Research staff acknowledged repeatedly at interview.

Largely as a result of having to cover a complicated literature, POLIS has a large and complicated field structure, which probably constitutes the most intimidating and off-putting aspect of the system (see Appendix 5 for a sample record). Through its field structure, POLIS offers the conventional search approaches - subject, publication form, author, title and date - and a slightly unconventional one - via Member - to its records. There is nothing really complicated or unusual about that, but the copious subject and publication form approaches offered do take some getting used to. Six fields contained subject information though it was possible to escape an indexing lesson on the individual characteristics and merits of these particular fields by using the default facility. No such help was afforded to the *form* searcher unless they sought refuge in the menu-driven version of POLIS. The form searcher - obviously thought by the system's designers to be found in large numbers in the House of Commons Library - had a big and bewildering range of choices placed in front of them. They could search on a **group** of publications (e.g. Parliamentary Questions), or they could search on an individual **form** of publication belonging to a form group (e.g. oral PQs, books). If they opted for the former, they were faced with another choice of nine form groups, and if they opted for the latter, they had 38 form groups at their disposal.

POLIS offered the traditional Boolean operators (**and**, **or**, and **not**) by which terms could be combined. POLIS does also provide for word proximity searching, though it is not called that (**Scan** is the command). It was, however, rather hidden from view of the searcher by both manual and menu, and is complex in execution. Through the **browse** command, it was possible to get access to the House of Commons subject thesaurus so that a term's controlled status could be checked: preferred/non-preferred

terms were listed and broader/narrower/related relationships indicated. Use of the on-line thesaurus involved exiting from the database that you were in and entering the thesaurus module: the procedure had to be repeated when returning to the database - the terms, which carry no postings with them, could not be carried with you. So the value of the on-line thesaurus was rather limited and there was no real advantage in using the on-line rather than the hard copy thesaurus. The **look** command, which provides for a display of the (inverted) field indexes was, however, available within the database being searched. The **look** command (surprisingly, not available in menu-driven POLIS) provides for an alphabetic examination of the words/phrases contained in POLIS database. Unfortunately, the logs provided only global data on the use of most of the commands mentioned in this paragraph: it was not possible to get a breakdown for end-users.

As most end-users searched POLIS in menu-driven mode, to fully understand their searching behaviour it is necessary to consider their design and construction, especially so since they suffered from many of the rigidities and confused choices that are commonly associated with menus. (A sample set of menus are reproduced in Table 4.14.) As with most menu-driven interfaces, this one guides and channels users' responses, making some searches very easy to make (usually- though not in POLIS' particular case - the well-trodden ones) and others (the unusual or different) more difficult to conduct. The choices provided by the menus and the framing of the questions obviously influence the search process. The proportion of Research Assistants that reported difficulties with the menus in the survey was significant but not unduly large: one in four respondents encountered problems when using the menu system. The major complaints were over the framing of the questions, some of the presumptions made and the limited range of options provided (section 4.3.4).

The shape and influence of the menus on searching behaviour will be returned to latter on in the specific discussions of searching behaviour (Section 4.3.3.4), but a few points bare mentioning now. Firstly, the opening menu encourages the end-user not to search the system in command driven mode - *experienced users only* it said. Though what end-users made of the word *freely* (see greeting screen, Table 4.14) was far from clear. Secondly, the main searching menu 1 attempted to get users to formulate their searches with form of material principally in mind. Thirdly, users were always reminded of the presence of archival files (Menu2.1). Finally, the menus were not always clear or well explained, and at no point did they tell the user that help was available.

4.3.2 The POLIS users

During the survey period, a total of 90 POLIS passwords were active, though in fact the number of POLIS users was much higher than that, for passwords were generally allocated to terminals and in some cases a number of users shared a terminal. In fact, there were a few instances where a dozen or so people shared the same password, and in one notable instance - that of Research Assistants using the Branch Library - perhaps as many as a hundred did so. In fact, at the time nobody knew for certain the total size of the POLIS user group, though it was thought to be in the range of 200-300.

POLIS was used by a wide variety of individuals - not all of whom were members of the House of Commons, and for a large number of purposes - not all of which related directly to information seeking. In consequence, any use of POLIS resulting from the maintenance of the database and anyone other than Members, their Research Assistants and House of Commons Library staff, were excluded from the analysis. Specifically, this meant excluding the use made of POLIS by: Library cataloguers, indexers and secretaries; House of Commons administrative staff (e.g. Table Office, *Hansard* Reporters Room), House of Lords staff and external organisations (other national legislatures). A little indexing use might have leaked into the analysis, as a number of front line Library staff, for instance International Affairs staff, indexed as well as searched. In theory, they should all have used an indexing password when indexing but in practice, on a few occasions they may have forgotten to do so. This meant attention could be given to just two (apparently) distinct groups of users: Members and their political staff (in practice, largely Research Assistants) and the House of Commons Library Staff.

4.3.2.1 The End-users.

As mentioned earlier it was Members' Research Assistants and not Members who did all the POLIS searching at The House of Commons. It was not possible to be sure that Members did not search POLIS during the survey period, but it was highly unlikely. No librarian or Research Assistant could identify any MPs as having searched POLIS, nor did any of the MPs interviewed admit to doing so - in fact, some had not even heard of POLIS by repute. This, of course, was highly significant in itself for it would be difficult to think of another group which delegated the on-line search so

blindly and comprehensively.

As a result, the POLIS end-user survey featured a group that operated as intermediaries too, although none possessed library qualifications and most were aspiring politicians or minor politicians in their own right, writing the speeches and briefing papers for their Members. The Research Assistants surveyed formed a largely young (over two-thirds were under thirty) and shifting population (80% had been working at The House less than three years).

For monitoring purposes it proved possible - and useful - to distinguish between two types of Research Assistant - those that used the Branch Library terminal, which was for the exclusive use of researchers, and those that dialled into POLIS from their House of Commons or constituency offices. The former group numbered - potentially anyway - some 600 and were represented by a single password - all searches were done at the same terminal, whereas the second group of Assistants numbered some 22 and was largely formed of Assistants to Opposition Members with constituencies outside of London, who had problems accessing the Branch Library terminal. Of the 60 Research Assistants questioned 57 (95%) searched POLIS, showing that if MPs did not use it their staff certainly did. All most Research Assistants obtained in the way of training was a five minute demonstration of the menu-driven system as part of the Library tour, although a small number (6) had received training in the (original) command-driven version. Compared to the frequent and extensive training Library staff obtained, Research Assistants received very meagre fare: indeed, three quarters of them felt they had not been trained in the use of POLIS at all. There was an irony in all this, for given their unfamiliarity with on-line services, the high turn over of staff and the lesser opportunities the had to search, it was the end-users who really needed the training

The two groups of on-line users - there was some overlap in membership - differed in some important ways. Firstly, and most importantly, the dial-up users were constrained, in the 'normal' on-line way, by cost (in their case, only by telecommunication costs: no connect charges were levied). This, together with the fact that their terminals - personal computers, were used for other purposes such as word processing, meant that users were more likely to keep searches short(er) and to log off after every search. What this group did not have to contend with though was the long walk and ever-present queues that the Branch Library users had to put up with. The Branch Library terminal, by contrast, was a single-purpose station, which was in

constant use and frequently left on from first switch on in the morning to close down at night. What all this meant to the analysis was that the data obtained on connect time and number of search sessions conducted were going to be more realistic for the dial-up group. Secondly, the Members that the dial-up users represented were thought to be generally highly motivated individuals (although just a letter was required to obtain access) and their staff too were thought to form a more motivated and dedicated group of users.

Because it was the Branch Library end-users who accessed POLIS in a very similar manner to the Library staff - through in-house dedicated terminals, their searching behaviour viz. a viz. Library staff was of particular interest and obtains special attention in the following analysis.

As mentioned earlier, Research Assistants largely used POLIS in its menu-driven form. Although the exact proportion of searches conducted in menu-driven mode was not known, what was known was that: (1) users of the Branch Library terminal were requested not to take the system out of menu-driven mode (a note was placed by the terminal requesting this - and there was the warning posted on the POLIS greetings screen stating that the command-driven option was only for experienced users); (2) only a small proportion of researchers - the questionnaire results suggested about 10% - were familiar with the command-driven mode and some of these were trained on an earlier variant, POLIS 1, which was a quite different system; (3) while the Branch Library terminal was left almost permanently in menu-driven mode, dial-up users were automatically connected to the command-driven POLIS and had to exit from it to enter the menus - this makes this group even more interesting from the point of view of the analysis to come, for it was likely that some would have *experimented* with the command-driven form.

4.3.2.2 The Library Staff

All staff who had direct contact with Members or their personal staff (about 87 people), plus anybody else who needed it (indexers for instance), were trained in the use of POLIS. But really out of a total staff of 150+ it was only the secretaries, attendants and Vote Office staff who were not trained in the use of POLIS. In an attempt to isolate differences in on-line searching practices amongst Library staff, resulting from differences in job responsibilities, subject affiliation and information retrieval training, this group was divided and sub-divided in to smaller and more homogeneous user groups. The fact that a number of the Library staff were not in fact

professional librarians, but subject specialists gave this sub-analysis special interest, for in some respects the Library too contained a group of searchers who could claim some of the characteristics of end-users. This also enabled comparisons to be made between end-users and specific groups of information professionals. The Commons then provided a rainbow-like spectrum of on-line user groups to investigate.

The primary Library staff analysis was then into: staff of the Research Division (subject specialists, plus supporting information professionals); Reference Library staff of the Parliamentary Division (professional librarians and library assistants); and Public Information Office (PIO) staff (professional librarians and library assistants, largely operating a telephone-based service to the public). Because of the telephone nature of much of the work of the PIO, comparisons between them and the dial-up Research Assistants were of particular interest and are featured in many of the analyses.

The Library Research staff, who numbered 37 at the time, were further broken down by their subject briefs: Economic Affairs; Education and Social Services; Home Affairs; Science and Environment; Statistics; and International Affairs. This meant that another on-line searching variable - subject affiliation - could be examined. An analysis of their joint use of the POLIS terminal in the Members' Library (C Room) was also provided: the various Research sections' staff took turns to man the Research desk in C Room. The job of Library Research staff was to deal with Members' written and oral enquiries. They also provided over-the-phone advice and briefing notes. Most of the Library Research staff were not professional librarians, but highly qualified subject specialists, usually recruited through the Civil Service Commission. In on-line terms, they would be best described as academic subject specialists and were thought to be on a scale somewhere between professional Library staff and Members' Research Assistants in terms of on-line competence. One or two professional librarians were attached to each subject section to assist in the maintenance of each section's information holdings. A complicating factor in the analysis of Research staff's use of POLIS was that sometimes professional librarians attached to the individual Research sections searched POLIS on behalf of the researchers (so some of the use of the Library research terminals resulted from a query being delegated twice: from Member to Researcher and from Researcher to Librarian). In the grey area between information professionals and end-users, the on-line performance and characteristics of the Library Researchers was of special concern throughout this investigation. For instance, to whom did they look for their searching styles - professional librarians or end-users?

The 22 strong Reference Library staff were further sub-divided into three groups: those who manned the Oriel Room of the Members' Library, the principal source of Parliamentary material for MPs (this is where they came if they wanted to trace a debate or question of the House, if they wanted to find out what stage a Bill has reached or if they required a Parliamentary Paper); those that staffed the Reference Room of the Members' Library, and who undertook work similar to that found in any Public Library reference department (quick reference enquiries in the field of current affairs were the unit's speciality); and Branch Library staff, who largely dealt with queries from Members' Research Assistants. The Branch Library staff were also members of the PIO, where they provided a telephone reference facility for Members and their Assistants and members of the public- some two-thirds of their work concerned the latter group. Given the different composition of the PIO's client group, and their largely telephone-based work it might be expected that their use of on-line might be somewhat different from their Library colleagues, and, possibly, there might be some similarities with the other telephone-based on-line group - the dial-up Research Assistants. Because of the pressures of working in the Oriel Room - close to the Chamber and (harassed) MPs' first port of call - the Staff there were the most highly trained of all - all were given frequent POLIS refresher courses and all had done spells in the indexing unit, where they developed a deep insight into POLIS cataloguing and indexing practices - something that was thought to be essential for good POLIS searching. If MPs were at the extreme (soft) end of the on-line spectrum, then the staff of The Oriel Room were at the opposite (hard) end.

4.3.3 Search characteristics

Because of the sheer volume of the data presented here and the consequent size of the chapter, an attempt has been made to make to the data more digestible. After every searching characteristic has been analysed, a summary has been provided, highlighting the salient points and making the key comparisons. These key summaries are placed in italics. In addition, all tables are structured along standard lines: comparisons are provided between: individual end-users groups - in-house (Branch Library) and dial-up; end-users as a whole and Library staff as a whole; individual divisions of the Library - Research, Reference, Public Information Office; individual units belonging to these divisions (Economics, Education, Home Affairs, International Affairs, Science Statistics, C Room, Oriel Room, Reference Room, Branch Library).

4.3.3.1 Use data

4.3.3.1.1 Frequency of use

According to the survey data a very high proportion (95%) of the 57 respondents used POLIS (Table 4.1). The fact that this was a sample survey, meant that actual searching numbers were undoubtedly higher than that, but given the fact that the sampling frame used was biased towards users, the proportion of users must be something less than that. Searching was highly variable though. Few people (11%) claimed to have searched it daily, though over half the sample (52%) said that they used it at least once a week or more frequently. A significant proportion (26%) used POLIS either irregularly or occasionally. If justification was needed for the introduction of the menu-driven system then this last figure probably provides it. Age seemed to be significant in determining whether the system was used or not: 63% of the researchers in the age group 18-30 searched POLIS at least once a week; 36% of those in the 31-40 category did so, but none of the over forties searched POLIS that frequently.

TABLE 4.1
FREQUENCY OF USING POLIS BY RESEARCH ASSISTANTS
:ANALYSIS BY AGE (Survey)

	18-30		31-40		40+		Total	
	n	%	n	%	n	%	n	%
Daily	5	12.2	1	9.1	0	0	6	10.5
More than once a week	13	31.7	2	18.2	0	0	15	26.3
Weekly	8	19.5	1	9.1	0	0	9	15.8
Fortnightly	4	9.8	1	9.1	1	20.0	6	10.5
Monthly	4	9.8	1	9.1	1	20.0	6	10.5
Irregularly	5	12.2	3	27.3	2	40.0	10	17.5
Occasionally	2	4.9	2	18.2	1	20.0	5	8.8
Total	41	100	11	100	5	100	57	100

NB Three respondents said they did not use POLIS.

The POLIS transactional logs were the principle source of use data, offering a range of statistical indicators: (a) the number of search sessions conducted; (b) the length of time spent on-line; (c) the number of records displayed; and, (d) the number of on-line transactions.

4.3.3.1.2 Number of on-line sessions conducted

Determining the number of on-line sessions conducted by users at The Commons was not easy, because many of the in-house POLIS terminals were left on all day, with little attempt being made to log off from the system at the end of a search. This was

largely because of convenience, but also because of the frequency and urgency with which searches had to be conducted, especially in the Oriel Room, where Members frequently required answers to their questions in a matter of minutes, if not seconds (speed of delivery is a key component of politicians' information needs). The half a minute it would take the system to load up would prove too much of a handicap (and embarrassment) for the librarian, and, at any rate, some terminals - like the Researcher's terminal in the Branch Library - were often in continuous use. The POLIS screen did go blank after 15 minutes of inactivity, but as far as the monitoring software was concerned, the user was still on-line.

Even so, a session generally did not stretch from switch on in the morning to switch off at night - some disciplined users did switch on/switch off at lunch time and, there were a number of other circumstances in which a session was terminated. Thus a session was terminated when a user: (1) changed database, that is, either moved to another POLIS file, or to an external host, which was accessed through the same terminal (the latter only applied to Library staff terminals); (2) renumbered their search or started the search again using the /n facility or the menu equivalent (LOGOUT); (3) examined the on-line thesaurus (using the **browse** command) or alphabetical index (using the **look** command). Thus, as an indicator of use, session data should be treated with some caution, though clearly it does denote some on-line activity and comparisons are especially valid. Also, as we shall see later, some of the other use indices, such as connect time, appear to lend support to the session data. Session data is therefore best taken together with other 'use' indices and best employed relatively (i.e. in inter-group comparisons). In the case of the dial-up users, who had every reason to log off from the system after completion of a search, the session data can probably be wholly relied upon to provide an accurate picture of the number of search sessions conducted. Indeed, if the pressure on the Research Assistant's Branch Library terminal was as great as Library staff said, then session times there were probably realistic, too. To overcome some of the problems inherent in the session data, hours and number of documents displayed have been sometimes used as the principle unit of division in search characteristic calculations.

Of the total of POLIS sessions conducted just 1307 (10.5%) were conducted by MPs' Research Assistants - an illustration of the largely delegated model of information seeking that was such a feature at Westminster (Table 4.2, column 1). The figure of 1307 sessions conducted cannot be a true portrayal of POLIS's potential worth for Research Assistants, for several hundred researchers were constrained by having to share a single terminal - in many cases a long distance from their offices. Even so

Research Assistants recorded more search sessions than all the Library units, with the exception of the Oriel room and the PIO. Branch library end-users (703) conducted a similar number of sessions to Reference Room staff (704). A more accurate picture of sessions conducted can probably be provided by the dial-up users: they conducted 604 search sessions - over 45% of the end-user total. With twenty-two dial-up users active during the period of the survey, it meant that, on average, each user conducted 27 on-line sessions during the year, which if the questionnaire survey (Table 4.1) was accurate meant that they used the system less frequently than their in-house colleagues.

TABLE 4.2
USE OF POLIS: SESSIONS CONDUCTED, CONNECT TIME
AND RECORDS DISPLAYED (Logs)

	Number of sessions conducted	Total time on-line (hours)	Ave. length of session (mins.)	Number of records displayed	Number displayed per session
END-USERS	1307	1173	54.1	96804	74
Branch library	703	1075	91.7	80499	115
Dial-up	604	103	10.2	16305	27
LIBRARY STAFF	11161	23007	123.7	1570057	141
Research	4111	9635	140.6	453814	110
Economics	566	1171	124.1	26357	47
Education	401	1165	174.3	43781	109
Home Affairs	222	695	187.8	18085	81
International Affs.	1248	2782	133.8	79760	64
Science	363	797	131.8	59859	165
Statistics	993	1714	103.6	136415	137
C Room	318	1310	247.2	89557	282
Reference	4120	9707	141.4	741814	180
Oriel Room	2489	6191	149.2	594829	239
Reference Room	741	2179	176.4	37840	51
Branch Library	890	1337	90.2	109145	123
Public Inf. Office	2930	3665	75.0	374429	128
TOTAL/AVE.	12468	24185	122.5	1666861	137

Comparison with Library staff

Folk wisdom in the Library had it that staff were divided in their attitude towards POLIS, with Research staff being more hard-copy oriented - and as a result somewhat anti-POLIS, and Reference staff strongly pro-POLIS. An initial reading of Table 4.2 does not appear to bare this out, for both groups made equal use of the system according to the sessions data. However, this is a little misleading for there were many more Research staff than *counter* Library staff, and Research staff had more POLIS passwords. It should also be remembered in all comparisons between professional librarians and subject specialists (researchers), that, sometimes, it was the

Library Executives who did the searching on behalf of the Research staff - making clear-cut comparisons difficult.

Looking more closely at the performance of individual research and reference units, it is clear that there were very big and very small users of POLIS. Not unexpectedly, the biggest user - The Oriel Room - had the most terminals (three). Oriel staff, around whose needs POLIS was largely designed, not surprisingly, conducted the most on-line sessions (2500). On the basis of a 240 day working year, that was over ten sessions per day. POLIS was, after all, an essentially Parliamentary material database, so it was probably inevitable that the room with the Parliamentary brief and major collection should feature so highly in its use. This was further brought home by the fact that, the Reference Room, despite having two terminals conducted barely a third the number of on-line sessions. Use then was not simply a function of access. The Reference Room dealt with quick reference enquiries, which could be answered with resort to reference works, cuttings files and current affairs databases (particularly FT PROFILE) - so its dependence on POLIS was naturally not so high. Significantly, too, the Oriel Room was the first room Members entered - and probably their first port of call.

Of the research units, International Affairs clearly stuck out - their use of POLIS stands head and shoulders above that of the other groups. Putting their figures in perspective: they conducted 26% ˅more sessions than Statistics, the next heaviest research unit, and 462% more than Home Affairs, the smallest user of them all. International Affairs was a busy and big section - it had more Library support staff than any other unit - and performed both research and reference functions - they were given two POLIS terminals for that reason. More pertinently perhaps, International Affairs took a conscious decision to reduce cuttings and to rely more heavily upon on-line databases - and it plainly showed. Conversely, Home Affairs had a reputation for being antagonistic towards POLIS - and despite its high research staff complement of five - it was POLIS' smallest user. Returning to International Affairs, their high session figures can also be attributed to their high use of PROFILE for this involved them signing off and on from POLIS quite a lot. Statistics had one of the biggest workloads of all the sections and was also, by common consent, the most computer-friendly of all the sections; their high use of POLIS bore witness to this. Science's relatively low session figures was possibly explained by its huge cuttings files - once the biggest of all - once the biggest of all the sections, but much reduced in size - and the lingering preference for hard copy on the part of some of its research staff. Perhaps too, its subject brief did not lend itself to POLIS use for science was not well

represented in parliamentary publications. Education's low usage can be attributed to the fact that it was another Library section which preferred hard copy and had a tendency to marginalise on-line databases. The attitude that prevailed in the section was encapsulated by this research staff member's comment: *I use POLIS infrequently; it is used as a last resort, usually when time is at a premium.*

On the evidence of session data Library staff were patently the biggest users of the system, but then they did have much better access to the system, and certainly had greater awareness and more training. On paper end-users had more terminals connected to the system - 23 as against 18 for the Library staff, but set against this was the fact that 22 of these were dial-up and multi-purpose work stations (with the inevitable restrictions that implied,) and the other station - the Branch library open access one - was at quite a distance from where most Research Assistants worked and shared by 100 or so researchers). In addition, the Library staff's figures were boosted by the fact that they had more databases they could switch to - something which inflated the session figures. As to the other comparison between the professional librarians and Library research staff: if use was simply a function of the number of staff and terminals then it might have been expected that Library researchers would have conducted the most sessions (they had more staff and terminals), but in reality they conducted less sessions. The fact that just the 7 staff of the Oriel room recorded nearly two-thirds as many sessions as the 37 members of the Research staff, suggests that the use of a database is primarily, a function of the relevance and immediate worth of the database. However, cultural and political factors would seem to have an impact.

4.3.3.1.3 Connect time

For the same reasons as furnished at the beginning of the session section, the figures for the amount of time spent on-line during a session at the Branch Library terminal were equally problematical. Essentially what could not be assumed was that searches were always being conducted whilst the terminal was registered as being on-line. This can lead to what appears to be very high connect times. Given that the system tends to be parked in the most current and heavily used file (CURRENT), a better idea of session connect times might be obtained by looking at search sessions conducted in other databases. Also dial-up figures, and the survey, are likely to provide more meaningful data.

Table 4.2 (Column 2) shows that Research Assistants spent 1178 hours on-line to POLIS in the year surveyed: an impressive figure when considered on its own, but one that soon pales by comparison with the 23007 hours registered by Library staff. You are talking here about 6-7 hours end-use per working day and an average of 25 hours per week, though with the very uneven searching patterns, a lot more than that in some weeks (Table 4.10). The disparity between Branch Library users (1075 hours on-line) and dial-up ones (103) was so great that the two sets of figures need to be considered separately. Plainly, dial-up sessions were conducted more quickly, for in respect to sessions conducted the two groups were more evenly matched. Overall, dial-up users were on-line for an average of about 4 hours 40 minutes on-line - not really sufficient time to become really proficient in the use of the system. However, in reality, (and as confirmed by the survey) just a relatively small proportion of users accounted for the majority of the use. The Branch Library's figure of 1075 hours made it one of the most lightly used terminals in the House of Commons.

As mentioned, House of Commons Library staff spent 23,007 hours logged on to POLIS in one year - a huge figure, even allowing for the liberal manner in which the figures were compiled. This figure - reinforced by others to come - points to the central role that POLIS played in the work of the Library. The relative use made of POLIS by the three principal Library staff divisions changes with the introduction of the connect time use indicator. The main change concerns the Public Information Office, whose share of use slipped some 10 percentage points. This can be explained by the different approach to on-line searching necessitated by the telephone nature of much of their work, for their average connect time was only half (53%) as long as for the other two groups. Looking more closely at the performances of individual Library units, it is noticeable that the performance of the Oriel Room was even more dominant (by virtue perhaps of its longer service hours - Library staff frequently worked into the night). Connect time figures provide even more proof of the weak position that POLIS holds in the work of the Home Affairs and Science Sections. The POLIS terminal used by Research staff manning their desks in C Room would, from the figures presented in Table 4.2, appear to be left on for long periods of time.

Session length

61% of the respondents in the survey said their searches were conducted within 10 minutes (Table 4.3). While this was not corroborated by the general end-user figure (47%) in the logs (Table 4.4) - possibly because of their imprecision, the survey finding was in fact supported by the dial-up log figures, which showed that approaching two-thirds (63%) of searches were accomplished within 10 minutes.

POLIS is largely a locational device so it would not be expected that Assistants would spend a lot of time on the system.

TABLE 4.3
DURATION OF RESEARCH ASSISTANTS'
POLIS SEARCHES (Survey)

	n	%
1 - 5 minutes	10	17.5
6 - 10 minutes	25	43.9
11 - 20 minutes	14	24.6
21 - 30 minutes	2	3.5
Varies greatly	6	10.5
TOTAL	57	100

The logs indicated that end-user sessions lasted on average about 54 minutes, making their sessions by far the shortest in The Commons (Table 4.2, column 3). This broad statistic is rather misleading - though it does correspond almost exactly with that which Lancaster *et al* (1994) came up with, for there was a wide variation in the average connect times of Branch Library users (92 minutes) and dial-up users (10 minutes). Even so, if we just take the Branch library figure alone, Members' Research Assistants still spent less time on their searches than all but two Library sections - the PIO (75 minutes) and the Branch Library librarians (90 minutes). It could be that end-users simply conducted their searches more quickly than the Library staff, but other factors could have played their part. The menus could have had something to do with this, for most menus offered the choice of restarting a new search: it is often easier to do this than disentangle yourself from the menus, and if this choice is taken, the monitoring software counted this as the end of one session and the beginning of another - and for this reason, the Research assistant session and connect time figures were perhaps more realistic than those of the Library staff. Another factor that could have led to relatively shorter searching sessions was the fact that - in theory anyway - Research Assistants had a ten minute limit imposed on them by the Library: the aim of which was to prevent people from monopolising the single terminal, but this rule was seldom adhered to or enforced, though no doubt with queues forming regularly, there would be an element of self-policing at work. In some ways, it must be doubted whether anything worthwhile could be accomplished in that time given the lack of expertise of the researchers, the cumbersome nature of menu-driven searching, the slowness of printing facilities, yet dial-up end-users seemed to manage.

Library research and reference staff's search sessions lasted approximately the same amount of time - 2 hours 20 minutes. Those of the Public Information Office were much shorter - 1 hour 15 minutes on average; there were a number of factors at work here - their very heavy enquiry load, a relatively shorter working day (they are not open in the evenings) and also the amount of database switching that they did (Table 4.11).

Possibly, by examining the connect times for the POLIS retrospective files, House of Lords database (HOLLIS) and EDM (Table 4.12), a better idea of how long it actually takes to do a search can be obtained. In fact, in the case of these files, average end-user connect times were much shorter, varying from 13 minutes in EDM to 26 minutes in HOLLIS. Dial-up users were even quicker still - for instance, averaging 10 minutes in PARL83 compared to the Branch library's 22 minutes, but in this case, were no shorter than those conducted in CURRENT. On average Library staff's sessions in the other files were much the same length as those of the dial-up end-users. It must always remain a possibility that, because there was much more in CURRENT, users spent more time in it.

The spread of connect times was so wide that it is worth breaking them down by time band (Table 4.4). There was a polarisation between very short session times (32% were over in less than 5 minutes) and very long times (23% were of 60 minutes or more). Big differences emerged between Branch and dial-up end-users: Branch library sessions were much longer, with a preponderance of long search sessions - 43% were an hour plus. While no Library group exactly replicated the Branch Library set of times, there were strong similarities with the most professionally trained of all sections - the Oriel Room.

<div style="text-align:center">

TABLE 4.4
DURATION OF POLIS SEARCH SESSIONS (Logs)

</div>

	mins.	0 - 4	5 - 9	10 -19	20 - 29	30 -59	60+
END-USERS	**n**	**421**	**187**	**193**	**94**	**189**	**303**
	%	**32.2**	**14.3**	**14.7**	**7.2**	**8.3**	**23.2**
Branch Library	n	155	70	54	45	77	302
	%	22.0	10.0	7.7	6.4	11.0	43.0
Dial-up	n	266	117	139	49	32	1
	%	44.0	19.4	23.0	8.1	5.3	0.2
LIBRARY STAFF	**n**	**3103**	**729**	**716**	**508**	**1029**	**5076**
	%	**27.8**	**6.5**	**6.4**	**4.6**	**9.2**	**45.5**
Research	n	964	233	252	183	370	2109
	%	23.4	5.7	6.1	4.5	9.0	51.3
Economics	n	112	34	42	32	55	291
	%	19.8	6.0	7.4	5.7	9.7	51.4
Education	n	91	20	18	10	27	235
	%	22.7	5.0	4.5	2.5	6.7	58.6
Home Affairs	n	59	15	7	12	12	117
	%	26.6	6.8	3.2	5.4	5.4	52.7
Internat. Affairs.	n	242	64	79	57	137	669
	%	19.4	5.1	6.3	4.6	11.0	53.6
Science	n	76	35	23	18	32	179
	%	20.9	9.6	6.3	5.0	8.8	49.3
Statistics	n	311	54	78	51	99	400
	%	31.3	5.4	7.9	5.1	10.0	40.3
C Room	n	73	11	5	3	8	218
		23.0	3.5	1.6	0.9	2.5	68.6
Reference	n	1041	284	262	173	392	1968
	%	25.3	6.8	6.4	4.2	9.5	47.8
Oriel room	n	633	196	174	100	202	1184
	%	25.4	7.9	7.0	4.0	8.1	47.6
Reference room.	n	99	28	41	37	92	444
	%	13.4	3.8	5.5	5.0	12.4	59.9
Branch library	n	309	60	47	36	98	340
	%	34.7	6.7	5.3	4.0	11.0	38.2
Public Info. Office	n	1098	212	202	152	267	999
	%	37.5	7.2	6.9	5.2	9.1	34.1

The figures for the dial-up group bore little relationship with the figures for any of the in-house groups. Thus in complete contrast to the large proportion of long search sessions conducted by their Branch Library colleagues and Library staff groups - typically 50% of their searches were over an hour in length, less than 1% of dial-up searches were of this duration. This pattern of session times for the dial-up users was in fact similar to that for TEXTLINE (Table 4.29). If it is assumed that the dial-up data was much more indicative of MPs' researchers searching behaviour, then this data can be used to obtain a better estimate of the number of on-line sessions conducted on POLIS by Research Assistants in the Branch Library. Thus, if it is assumed that a typical search session lasted ten minutes, then with 1075 hours spent on-line, the number of actual search sessions conducted would be more likely to be 6450 rather than 1300. However, this is to assume that in-house and dial-in use was of a similar type - and data elsewhere, and the writer Sloan (1992), suggest that they may well be different. Given the reliability of the dial-up figures, additional data for

the period October 1987 - when dial-up facilities for Members were first introduced - to August 1988 was collected. During this eleven month period, 233 search sessions were undertaken, and 38 hours spent on-line - showing that dial-up use was on the increase. The length of search sessions remained very much the same - an average of about 10 minutes.

Comparison with Library staff

For Library staff a large proportion (around 50%) of sessions lasted more than an hour, but quick (less than 5 minutes) searches were also quite common The latter cannot simply be put down to in and out searches of archival files, for use of these files accounted for only 2.5% of all searches. Maybe, Library staff, either because they had tidy minds, or because they found it less confusing if searches were renumbered at the start of a new on-line session, did use the POLIS /**n** facility quite a lot - if this was true it does of course lend more credibility to the session figures. The differences between Librarians and Library Researchers was not great, though Researcher's sessions were typically longer. The PIO figures deserves special mention, for nearly 38% of their searches were completed in less than 5 minutes. Now this too is very gratifying information for this adds even greater credibility to the session figures as we know that much of their searching was done in connection with a telephone enquiry, and they did aim to provide an answer without returning the call (something which was bound to speed the search process).

While the end-users of the Branch Library terminal had a distinct spread of search session times, they were not that dissimilar to several of the library staff user groups - most notably that of the Oriel Room. Indeed, there were greater differences amongst individual library groups than between librarians and end-users. However, none of the Library sections had distributions remotely similar to that for the dial-up end-users, though interestingly, the PIO, which provided a largely telephone-based service came the closest. Unquestionably dial-up users completed their searches far more quickly, but we cannot say that about end-users in general. Method or form of access was obviously a major influence on searching behaviour.

4.3.3.1 4 Records Displayed

The POLIS monitoring data supplied figures on the number of records displayed per search session, which was not quite the same thing (but could be) as the number of postings obtained by a search. These records might also have been printed out, but not

necessarily so. It probably can be assumed though, that the number of records displayed provided the best evidence so far of the use of the system because, while the monitoring software could not always distinguish between an active search session and a period when the system was on-line, but at rest, the display of documents was a conscious, demonstrable act.

Research Assistants showed how much they relied on POLIS by displaying a sizeable 96,804 records in a year: an average of 74 items displayed or printed out per search session (Table 4.2, columns 4/5). Again, the average figure tends to hide things, for there were large differences between Branch Library and dial-up users, and, there was a wide range of practices even within these two groups.

Taking the Branch Library users first, what perhaps is most noticeable with this group was the very large number of items that were displayed. On average 115 documents were displayed per session. More than one-quarter of their search sessions resulted in display runs of a 100 or more items (Table 4.5): a finding which would not surprise Branch Library staff, who have alluded to such practises in interviews with them. Whether such large displays were due to: sloppy techniques (asking for a big print run without realising it and leaving the records to scroll over the screen); extremely broad searching yielding lots of data (and that too has been observed by Branch Library staff); the inevitable result of long sessions; or, indeed, were what was genuinely required (comprehensiveness), is not certain. Probably a mixture of all these factors. What ever, Research Assistants were in good company for Library staff displayed on average even more records per session (141) and Oriel Room staff had proportionally more sessions which displayed more than a hundred records (33%).

Turning to dial-up users now, the Table does show that, in total contrast to their in-house colleagues, very few (under 6%) of their sessions ended up with document displays of 100 items or more. Most frequently, sessions resulted in 1-9 items being displayed 27% of all searches ended up with a result in this range.

Comparison with Library Staff

House of Commons Library staff displayed a spectacular one and a half million records in the year surveyed. It is, of course, very difficult to take this massive statistic in. Perhaps it is best put into perspective by relating it to the number of users on whose behalf the data would have been displayed, if not, passed. Thus, on average each Member had approximately 2400 records displayed on their behalf. Real

differences between the three main Library staff groups opened up when use is measured by the number of records displayed. Library Reference staff, whose use of POLIS up until now had been almost identical to that of the Research staff, were by this criteria clearly the biggest users of the system. They accounted for something approaching half (47%) of all the POLIS records displayed. On the basis of sessions conducted and connect time, they were within a percentage point of the Library Research staff, but when on-line activity was measured in records displayed, they were almost 20% ahead. This hints strongly at a difference in searching styles. An average Reference staff's on-line session yielded 180 records, whereas the figure for Library Research staff was 110. Intuitively, one might have thought the very opposite would have occurred, for Reference Library staff were more partial to controlled vocabulary searching (Table 4.16) and that probably should have led to a higher degree of precision. Again, the individual Library sections demonstrated a wide range of practices. There really was no norm for there were the really big displayers (C Room, average 282 per session; and the Oriel Room, 239) and the very small displayers (Reference Room, 51; and Economics, 47). The differences in practice were huge, pointing to the fact that, even amongst a group as apparently homogeneous as Library staff, searching outcomes can vary enormously. The large number of records generated by Statistics - 136415, the highest of all the research sections, clearly stood out. This can be explained by the fact that, in an attempt to anticipate the kinds of questions they were going to be asked, Statistics staff did a lot of on-line scanning of the description field of recent Parliamentary Questions (a rare case of on-line browsing).

What the display figures do confirm is the position of the Oriel Room as the busiest POLIS user group: on the four measures to date - sessions conducted, connect time and items displayed - they record the highest figures. Suggesting that there must be some correlation between the various use indices. They scored most impressively, though, on the last measure: the 594,829 records they displayed represented 38% of all the records displayed by the House of Commons Library staff (and they had only 17% of the Library searching staff's terminals).

Library staff sessions, like those for end-users, tended to either produce lots of records (23% produced a hundred or more) or very few (40% produce less than 10), but then sessions also tended to polarise in terms of connect time into long and short ones. However, it does not seem that display figures were a function of the length of on-line sessions - see page 98. What, one wonders, causes so few search sessions to generate between 10-19 items (10%) and 50-99 items (12%).

Zero hits

Nearly one-quarter (24%) of end-user sessions ended in the display of no items. Significantly, this figure was not that far removed from the proportion (16%) of questionnaire respondents who said that they were either, only sometimes satisfied or not satisfied with their POLIS searches (Table 4.19). By comparison with the House of Commons Library staff, 18% of whose searches ended with nothing, Research Assistants did experience more zero hit sessions. Branch Library users experienced less zero hits sessions than their dial-up colleagues - though only marginally so (2.5% less). In fact, it was the dial-up Assistants who produced the second largest proportion of nil result searches of all the House of Commons on-line groups: 25.5% of their searches failed to display any records - they were beaten for first place by the Science Researchers, whose every third search ended with nothing. Perhaps the dial-up users were not the proficient searchers the Library staff thought, or more likely, they just did not have to hand the professional help in the shape of Branch Library staff. Reference Library staff and Library researchers registered the same proportion of zero hits, so no obvious differences in skill are evident here then. Of the big three searching divisions it was the PIO, which recorded the highest proportion of nil searches - 21% in their case. Perhaps, they were suffering the higher failure rates associated with a telephone based, public oriented operation. Delving deeper among the sub-groups, it was, appropriately, The Oriel Room staff - the Library's most highly trained POLIS searchers and the Statistical section, the Library's most computer literate group, who recorded the lowest proportion of zero hit searches.

Of the individual Library Research and Reference units, the Reference Room generated a large proportion (29%) of no displays. Fittingly, the work of the Reference Room would appear to be of the quick reference kind, with an additional 28% of sessions producing between 1-9 records, and only 7% producing a hundred or more.

Can the zero hit figure really be represented as the user failure rate? Possibly, but clearly there are two situations when this would *not* necessarily be the case. Firstly, it has to be remembered that a zero score would be recorded if the system was at rest in CURRENT and another database was called up. What dents this argument though is the fact that the Reference room did not change POLIS files that much (Table 4.11). Secondly, negative searches can prove to be positive in the case of POLIS - take for example the situation where a Member requires confirmation of whether a PQ they wanted to table did not breach the House rule, that the same question cannot be tabled within three months - in which case a no hits search produces a positive result.

Nevertheless, even allowing for this, there were still a lot of searches that found nothing, so the figure needs to be treated with caution. What lends credibility to the measure as a possible indicator of poor performance is the fact that the Oriel Room, widely regarded as the Library's most proficient POLIS user (they get more training and more than enough practice), made proportionally the fewest nil searches of all on-line groups and Science, the lightest user, the most.

TABLE 4.5.
NUMBER OF POLIS RECORDS DISPLAYED PER SEARCH SESSION
(Logs)

		0	1 - 9	10 - 19	20 - 49	50 - 99	100+
END-USERS	n	309	266	155	215	147	214
	%	23.4	20.4	11.9	16.5	11.2	16.3
Branch Library.	n	158	101	71	105	87	181
	%	22.5	14.4	10.1	14.9	12.4	25.7
Dial-up	n	151	165	84	111		60
	%	25.0	27.3	13.9	18.4	9.9	5.5
LIBRARY STAFF	n	2048	2440	1132	1643	1363	2555
	%	18.3	21.9	10.1	14.7	12.0	22.9
Research	n	729	861	462	690	533	836
	%	17.7	20.9	11.2	16.8	13.0	20.3
Economics	n	130	124	70	98	69	75
	%	23.0	21.9	12.4	17.3	12.2	13.3
Education	n	60	101	50	61	57	72
	%	15.0	25.2	12.5	15.2	14.2	18.0
Home Affairs	n	40	61	42	31	26	22
	%	18.0	27.5	18.9	14.0	11.7	9.9
Internat. Affrs.	n	236	230	126	247	194	215
	%	18.9	18.4	10.1	19.8	15.5	17.2
Science	n	109	71	30	59	39	55
	%	30.0	19.6	8.3	16.3	10.7	15.2
Statistics	n	124	218	123	156	107	265
	%	12.5	22.0	12.4	15.7	10.8	26.7
C Room	n	30	56	21	38	41	132
	%	9.4	17.6	6.6	11.9	12.9	41.5
Reference	n	715	915	391	563	457	1079
	%	17.4	22.2	9.5	13.7	11.1	26.2
Oriel Room	n	324	508	207	340	299	811
	%	13.0	20.4	8.3	13.7	12.0	32.6
Reference Rm.	n	218	210	105	109	47	52
	%	29.4	28.3	14.2	14.7	6.3	7.0
Branch Lib.	n	173	197	79	114	111	216
	%	19.4	22.1	8.9	12.8	12.5	24.3
Public Info. Office	n	604	664	279	390	353	640
	%	20.6	22.7	9.5	13.3	12.0	21.8

Number of records displayed by duration of search

By dividing the number of records displayed by connect time a measure of productivity or search efficiency could be produced - and it does produce some interesting results (Table 4.6). Research Assistants did display more records per hour (82) than Library Staff (68), though this largely arises because of the influence of the dial-up end-users, who produced far more records per hour than any other on-line group (158) - and by a large margin. The in-house end-users recorded an hourly rate of 75 records, which is in the middle range of the rates for the individual Library units. In the particular case of end-users, statistical evidence was sought to determine whether long sessions (i.e. long connect times) were the ones that produced large numbers of items. As might have been expected, there was some correlation (i.e. the correlation coefficient was significantly greater than zero) but at 0.37 it was not very high overall. And this could well be because terminals were left connected, but not used. Supporting this argument, was the case of dial-up users, for them the correlation was much greater, ranging from 0.5 to 0.8 from month to month, with many short, productive searches but no polarisation at the other end (i.e. not many long, productive searches). Interestingly, Lancaster *et al* (1994) did not find any correlation either between long searches and the relevance of the documents retrieved. In the case of Library staff, a visual comparison of columns 4 and 6 Table 4.1 clearly shows that it was not a case of the longer the session, the more items displayed because, the sessions of The Oriel Room (big displayers) were on the shorter side whereas those of the Reference Room (short displayers) were on the long side - in fact, in these two cases, the very opposite was true, but it was not true in every case.

Of the Library groups The Oriel Room was easily the most productive with 96 items displayed per hour. The Reference Room produced the least number of records per hour. It is not clear why some units produced more records per hour than others, though it must be said the analysis was affected by the predisposition of users to leave the terminal on-line.

TABLE 4.6
NUMBER OF POLIS RECORDS DISPLAYED
PER CONNECT HOUR (logs)

END-USERS	82
Branch Library	75
Dial-up	158
LIBRARY STAFF	68
Research	47
Economics	23
Education	38
Home Affairs	26
International Affairs	29
Science	75
C Room	68
Statistics	80
Reference	76
Oriel Room	96
Reference Room	17
Branch Library	82
Public Inf. Office	102

Nearly a 100,000 records displayed in a year constitutes significant use of POLIS by Research Assistants. As Peters et al (1993) found, the real differences in display patterns were not so much between end-user and intermediary, but between the dial-up users and the in-house users. Indeed, the PIO's figures illustrated that there is much in common with telephone-based services of whatever kind. Dial-up searches were typically short and unproductive. It was true that in, general, end-user searches produced fewer records and more zero hits than those of the Library staff. But, as always, it was possible to find a group of Library searchers, who exceeded end-users in this. Plainly, zero hit searches are not just a phenomena associated with end-user searching. There is though, in the example of the Oriel Room, evidence to suggest that the most professional of user groups experience the least failures. Using records displayed per hour as a measure of productivity, it was the dial-up users who packed most in to the hour. There was no strong evidence to point to the fact that long searches produced the most records.

4.3.3.1.5 Number of on-line transactions

In the analysis of on-line use, a distinction can be made between a search session and

what is probably best described as an on-line interaction or transaction. Thus a single search session might involve a large number of individual interactions with the system: fields maybe selected, terms truncated or perhaps the index is examined to check on term availability. A search session might then consist of just one interaction or - more likely - many. The POLIS monitoring software furnished transaction statistics and these were, arguably, the most useful and important of all the monitoring data. The search commands monitored were as follows: the straightforward search (the **find** command on POLIS); the search conducted with truncation (*); the search on a range of items (-); the search which was automatically switched from a rejected form to a preferred form (a facility which existed for subjects, names of organisations and Members' names); field searches and the inspection of the (alphabetical) inverted index (the **look** command). Global data on the number of occurrences of these commands is provided here: more details about the individual search commands used can be found in Sections 4.3.3.4/5. Unlike the session data, which was complete, this data was generated on a sample basis, being furnished for the busiest day in the last ten working days of each month (except for December's data, which was inexplicably missing).

The number of transactions conducted can be a measure of busyness when on-line, and, indeed, a sign of proficiency also (showing a greater and wider command knowledge and more complex searching). Research Assistants made 1904 POLIS transactions on the 11 sample days surveyed: an average of 173 a day (Table 4.7). On a strictly one-to-one comparison with Library staff, they were outperformed, with the Librarians conducting over five times as many transactions. However, this placed them sixth overall in the list of on-line sub-groups, with 8 Library groups behind them - providing further confirmation of the heavy use of POLIS by Researcher Assistants. This meant that they were busier on-line than every Library Research unit, with the exception of International Affairs. Assistants using the Branch library facility were much more active than the dial-up users, accounting for 82% of all transactions, yet they only accounted for 54% of the on-line sessions, though they did spend much more time on-line.

TABLE 4.7
NUMBER OF POLIS TRANSACTIONS CONDUCTED
(Logs)

	n	%
END-USERS	**1904**	**100**
Branch Library	1569	82.4
Dial-up	335	17.6
LIBRARY STAFF	**10737**	**100**
Research	4700	43.8
Economics	425	4.0
Education	333	3.1
Home Affairs	195	1.8
International Affairs	2292	21.3
Science	186	1.7
Statistics	1033	9.6
C Room	236	2.2
Reference	3470	32.3
Oriel Room	2493	23.2
Reference Room	190	1.8
Branch Library	787	7.3
Public Information Office	2567	23.9

Comparison with Library staff

Clearly, in the light of on-line transactions, Library Research staff were POLIS' most active users, performing 4700 interactions on the eleven days; that was, an average of 427 a day. As Reference Library staff have proved busiest on account of sessions conducted, time on-line and records displayed, it would seem that Library Research staff might do less searching, but when searching they were more active: their searches were either more complex or convoluted - employing more commands and search terms and making more frequent changes in direction. A closer examination of the figures for the individual Library units shows that it was The Oriel Room (2493) and International Affairs (2292) who were the most active user groups, confirming their established positions as the high volume users. The Reference Room staff, already known as on-line users whose sessions produced the fewest records, were also the most economical or simplest searchers - they made only 190 interactions, compared to the 2493 made by their Oriel Room colleagues, next door: an enormous difference.

End-users were no where near as active in their use of search commands as the Librarians, but the activity gap closes when compared to sessions mounted: end-users

accounted for 15% of the transactions made, but only 10% of the sessions. The most proficient of the on-line groups, The Oriel Room was also the most active in the number of on-line transactions conducted, suggesting that the number of transactions might be a good indicator of search proficiency, too. Though you cannot rule out the fact that it could simply have been a function of heavy and regular use. However, other professionally trained groups did not fare so well - the Reference Room and the Branch Library for instance, and were less active than some of the Library Research groups - International Affairs and Statistics for instance.

4.3.3.2 Time of searching

The point of this analysis was to determine the time of day, day of the week and month, when on-line searching took place, and to determine whether there were any differences between end-users and Librarians in this. Of special interest was whether dial-up users availed themselves of the greater freedom to search POLIS.

Time of day

Table 4.8 shows the time of day when search sessions were undertaken. Because of the method of monitoring, the Table provides not just a record of when people logged onto the system, but also when they changed files and when they indicated that they were starting a new search session by using /n facility. So in many ways the figures contained in the Table provide an indication of levels of activity.

The main difference between the times Members' Research Assistants and Library staff did their searches was, that the latter did a lot of their searches in the morning (about 50% were undertaken before 1p.m.), and the former did most in the afternoon (55% after 1p.m.). The survey results had indicated that there was considerable bunching of on-line use during the day and, in fact, the Branch Library terminal was in pretty constant use from about 9 in the morning until 6 in the evening, with 10-12 and 2-4 periods being especially busy, accounting for 52% of all searches. The fact that there was only one terminal must inevitably flatten out use. Branch Library end-use time patterns mirrored most closely those of the PIO - another occasion when these two outwardly very different groups of on-line users had something in common. As expected, dial-up users were active for a much longer part of the day, their searching went on through the evening: nearly 10% of their searches were conducted then. No other user group had such a high proportion of evening searches. On-line databases obviously extended the information-seeking day of those Research Assistants. External users did not seem to start the day early though, with use not

peaking until 12-1 p.m.: Branch Library use peaked some two hours earlier.

TABLE 4.8
TIME OF DAY WHEN POLIS SEARCH SESSIONS WERE CONDUCTED
(Logs)

		-9 am	9-10	10-11	11-12	12-1	1-2	2-3	3-4	4-5	5-6	6-7	7 pm+
END-USERS	n	11	93	150	154	159	126	167	166	132	90	17	40
	%	1	7	11	12	12	10	13	13	10	7	1	3
Branch Library	n	9	58	99	72	71	65	96	98	86	49	-	-
	%	1	8	14	10	10	9	14	14	12	7	-	
Dial-up	n	2	35	51	82	88	61	71	70	46	41	17	40
	%	2	6	8	14	15	10	12	12	8	7	3	7
LIBRARY STAFF	n	254	1172	1960	1382	927	749	1136	1278	913	685	192	513
	%	2	11	18	12	8	7	10	12	8	6	2	5
Research	n	33	281	847	619	376	285	420	461	348	279	63	99
	%	1	7	21	15	9	7	10	11	9	7	2	2
Economics	n	4	31	91	81	68	40	79	69	48	49	1	5
	%	1	6	16	14	12	7	14	12	9	9	-	1
Education	n	5	33	84	85	46	37	24	29	35	18	2	3
	%	1	8	21	21	11	9	6	7	9	4	<1	1
Home Affairs	n	5	15	42	35	26	7	16	30	30	15	-	1
	%	2	7	19	16	12	3	7	14	14	7	-	1
Internat. Affrs.	n	4	84	273	182	92	70	128	126	90	89	35	75
	%	<1	8	22	15	7	6	10	10	7	7	3	6
Science	n	6	48	90	41	33	18	35	23	39	20	8	2
	%	2	13	25	11	9	5	10	6	11	6	2	1
Statistics	n	7	63	203	128	78	93	106	141	82	74	14	4
	%	1	7	21	13	8	10	11	14	8	7	1	-
C Room	n	2	7	64	67	33	20	32	43	24	14	3	9
	%	1	2	20	21	10	6	10	14	8	4	1	3
Reference		221	891	1113	763	551	464	716	817	565	406	129	414
		3	13	16	11	8	7	10	12	8	6	2	6
Oriel Room	n	101	340	374	259	154	136	192	208	171	149	99	306
	%	4	14	15	10	6	5	8	8	7	6	4	12
Reference Room	n	42	84	147	52	32	48	57	62	42	37	30	108
	%	5	11	20	7	4	6	8	8	6	5	4	15
Branch library	n	11	76	166	115	83	59	117	113	74	76	-	-
	%	1	9	19	13	9	7	13	13	8	9	-	-
PIO	n	67	391	426	337	282	221	350	434	278	144	-	-
	%	2	13	15	12	10	8	12	15	10	5	-	-

Comparison with Library staff

10 a.m. to 12 noon was obviously the busiest time for searching, for it was during this period when 30% of the on-line sessions were conducted. Of course, that was the time when Members were preparing themselves prior to entering the Chamber at 2p.m. Generally, only a little searching went on after 5 p.m. - 12% of search sessions took place after 5 p.m. Lunch time was relatively quiet, possibly because there were fewer Library staff available to search, and Members were also at lunch then. Individual Library Research units threw up some interesting characteristics. Science showed the greatest concentration of use, with 25% of their on-line sessions concentrated in just a single hour (10 a.m. to 11 a.m.). Education had a hectic two hours between 10 a.m. and 12 p.m., with 42% of their sessions starting then. Few Library Research staff undertook searches after 6, confirming the Library's 9-6 work

routine. It was really the Oriel Room and Reference Room, the key physical enquiry points, that searched on through the evening - the other units were largely closed.

Day of the week

Table 4.9 provides evidence as to the day of the week when searches were conducted. Mondays and Fridays were the quietest days - largely because Members were involved in travelling to and from their constituency on these days and, in the case of Monday, catching up on the weekend's sizeable post. Tuesday was by some margin the busiest day - Researchers gearing up for the week ahead? The downturn in use on Wednesday was a little puzzling. Perhaps it all had to do with the fact that there was no Prime Minister's Question Time on that day? - and Prime Minister Question Time always fills the House, bringing with it a flurry of information activity. There were only small differences between Research Assistants and Library staff - the basic differences being that: (1) Library staff spread the searching load more evenly throughout the week; (2) Dial-up users exercised their greater freedom to search by searching on weekends - 2% of their searches were conducted then.

TABLE 4.9
DAY OF THE WEEK WHEN POLIS SEARCHES WERE CONDUCTED[1]
(Logs)

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
END-USERS	17.1	24.8	19.5	22.6	15.5	<1	1
Branch Library	16.4	26.2	19.3	21.2	16.9	-	-
Dial-up	18.0	22.2	19.2	24.5	13.9	1	1.7
LIBRARY STAFF	18.8	21.8	22.6	21.2	15.4	<1	<1
Research	18.0	22.5	21.5	22.8	15.0	<1	<1
Reference	19.5	22.0	25.5	19.5	13.4	<1	<1
Public Information Off.	18.9	21.8	20.4	21.7	17.8	-	-

1 Expressed as a percentage

Month of the year

A monthly analysis and comparison of the three use indicators (Table 4.10) shows:

(1) Monthly values varied enormously (one month's figures could be as much as 7 times that of another month as in the case of dial-up users' sessions). The greatest variability occurred amongst dial-up users.

(2) Use - as judged by search sessions conducted - generally builds from Autumn, peaks in the Spring and falls right back in high Summer (the main Parliamentary recess).

(3) With the exception of sessions, there appears to be little correlation between the use figures for dial-up and Branch Library users. Thus, in the case of connect

time and records displayed one group's busiest month was the other groups lightest month, and vice versa. Take the case of connect time, it peaked in September for Branch Library users and was at its lowest in April, - yet, this was just the reverse for dial-up users. Thus, in the case of items displayed, the peak month for dial-up users was April - the low month for Branch Library users. Perhaps, what we were witnessing here was a migration from one form of access to the other, particularly during the recesses (dial-up users were also Branch Library users).

4) There was little consensus between indicators as to which was the busiest month: search session figures suggested February; connect time showed September to be the busiest month, and the number of records displayed peaked in February.

5) A lot of on-line searching was tied up with the business of the House, for when the House was not sitting (for example, in August and September) dial-up use was negligible. Branch Library use seemed to hold up reasonably well though: perhaps Library staff were taking advantage of the unused terminal?

TABLE 4.10
MONTHLY USE FIGURES FOR POLIS : END-USERS
(Logs)

Month	Search Sessions		Connect time		Records displayed	
	Branch	Dial-up	Branch	Dial-up	Branch	Dial-up
Sep	41 (6%)	10 (2%)	6680 (10%)	187 (3%)	7381 (9%)	208 (1%)
Oct	54 (8%)	41 (7%)	5664 (9%)	405 (7%)	3552 (4%)	406 (3%)
Nov	74 (11%)	29 (5%)	5456 (9%)	347 (6%)	8029 (10%)	701 (4%)
Dec	46 (7%)	42 (7%)	4806 (7%)	603 (10%)	3161 (4%)	2038 (13%)
Jan	48 (7%)	43 (7%)	5327 (8%)	479 (8%)	6595 (8%)	1868 (12%)
Feb	110 (16%)	75 (12%)	5004 (8%)	680 (11%)	15275 (19%)	1587 (10%)
Mar	69 (10%)	54 (9%)	6040 (9%)	450 (7%)	8642 (11%)	1654 (10%)
Apr	44 (6%)	70 (12%)	3949 (6%)	750 (12%)	2826 (4%)	2481 (15%)
May	55 (8%)	66 (11%)	4776 (7%)	631 (10%)	6052 (8%)	1765 (11%)
Jun	47 (7%)	63 (10%)	5688 (9%)	520 (8%)	3568 (4%)	1039 (6%)
Jul	59 (8%)	60 (10%)	5023 (8%)	559 (9%)	5939 (7%)	1407 (9%)
Aug	56 (8%)	51 (8%)	6114 (10%)	578 (9%)	9479 (12%)	1151 (7%)

End-user searching was highly volatile, swinging about dramatically from one month to another, and from one day to another. The rhythms were very much of those of The House: legislation and policies are announced in the Queen's Speech in the autumn, after which there is a general increase in work as bills are introduced and make their way through the House. (Bills have to be introduced and passed for Royal Assent in one Parliamentary session.). The busiest day of the week in The Commons - Prime Minister's Question Time - was also the busiest on-line day. Use was, however, evenly spread throughout the working day. Again, the pattern of dial-up use was different:

dial-up facilities seemed to be instrumental in spreading out the on-line searching load - throughout the day and week, though not the year. The pattern of Library staff searching was a little different, though not a lot. The principle difference being that their searching was less evenly spread through out the day and more evenly spread through out the week. All use measures, bar one, agreed on September as being the quietest on-line month and there was somewhat less agreement over the heaviest month - February.

4.3.3.3 File Selection and retrospective searching

During the study period, POLIS hosted 16 files, a good number of which were either experimental or in the process of development. Excluding the training file, there were five files which were strictly operational and available for searching. There were the three POLIS files - the (then) current file (CURRENT), covering parliamentary material from June 1987 onwards and all non-parliamentary material, and the two archival files - PARL83 (June 1983 to May 1987) and PARL79 (1979 to May 1983); Early Day Motion (EDM) file; and HOLLIS, the House of Lords pamphlet file (to April 1987, and which has since merged with the CURRENT file). File use can be examined in two ways: by counting the number of sessions conducted in each file, or by counting the amount of time spent in each file. It is sometimes thought that end-users are blinkered when they come to database selection, invariably using a single source, even when it was not the most appropriate: the analysis proceeds with this in mind.

Research Assistants used all the five files that were available to them, though the House of Lords file, which was not available to dial-up users, was only used 5 times (Table 4.11). Of the 1307 searches conducted by Research Assistants, 985 (76%) were undertaken in the CURRENT file. The second most popular file was the archival PARL83, in which 190 (15%) searches were made. While both sets of Research Assistants favoured CURRENT, the extent to which they relied on this one file differed. Dial-up users conducted most (87%) of their searches in CURRENT, but in-house Assistants used it nowhere near as much (66%). In-house users made greater use of the archival files. The difference could be partly put down to the fact that the menu-driven system constantly reminds the user of the presence of the archival files, and Branch Library users were most likely to use the menu-driven form. Indeed, two of the Research Assistants surveyed wanted more historical data on POLIS. They were unhappy that *Library materials go back only to 1985* (something which was in fact being slowly rectified by the Library). Interestingly, although Library staff were

paranoid about the (lack) of currency of the data on POLIS - especially in regard to Parliamentary Questions - none of the Assistants mentioned currency as a problem (typically data was entered onto POLIS a week after its receipt).

If the actual time spent in each file is considered (Table 4.12) a different picture emerges. The preference for CURRENT was now even more marked and the positions reversed - Branch Library researchers now spent (slightly) more time in CURRENT (92%) than dial-up users (89%). The reason for this was plain: Branch Library users' searches of CURRENT were longer than those conducted in any of the other files. Thus a session lasted 129 minutes in CURRENT (something of an exaggerated figure because the terminal rests in CURRENT), but only 4 minutes in EDM, 24 minutes in PARL79 and 22 minutes in PARL83. Puzzlingly, searches by dial-up users were longest in the EDM file (13 minutes): searches of CURRENT took only 10 minutes on average.

It is quite possible that the strong interest shown in CURRENT was attributable to something other than its currency, for after all it is the biggest POLIS file of all, accounting for some 51% of the records on POLIS. And while it contained just 14 months of parliamentary data at the beginning of the survey and 26 months at the end, it did contain non-parliamentary material going back to the early eighties (i.e. for some material, it is not that current).

TABLE 4. 11
POLIS FILE USE: NUMBER OF SESSIONS[1]
(Logs)

	Current	Parl83	Parl79	EDM	HOLLIS
END-USERS	75.8	14.6	8.3	0.9	0.4
Branch Library	65.9	20.8	11.8	0.4	0.7
Dial-up	87.0	7.3	4.1	1.5	-
LIBRARY STAFF	70.0	13.0	6.5	10.5	0.2
Research	80.5	13.1	5.2	1.1	0.2
Economics	81.3	11.1	3.5	3.7	0.4
Education	80.8	13.5	3.7	1.5	0.5
Home Affairs	83.3	12.2	3.6	0.9	-
International Affairs	89.9	6.2	3.9	-	-
Science	87.9	6.6	3.6	1.4	0.6
Statistics	64.7	24.8	10.1	0.5	-
C Room	80.8	14.8	2.8	1.6	-
Reference	68.8	12.1	6.4	12.5	0.2
Oriel Room	63.3	14.9	6.7	14.8	0.4
Reference Room	94.3	3.0	1.1	1.6	-
Branch Library	62.8	12.1	9.9	15.1	0.1
Public Information Off.	56.2	14.3	8.4	21.1	-

1. expressed as a %

Comparison with Library staff

Taking sessions first, the current POLIS file was by far the most heavily used, with 70% of all sessions conducted in it. The overriding concern of Library staff with current data becomes even more obvious when calculations of current file use are made as a percentage of the use of all three POLIS parliamentary files - the figure is 78% (the equivalent figure for end-users being 75%). The Library Research group was most preoccupied with CURRENT - 81% of the sessions conducted in the three POLIS parliamentary files were conducted in CURRENT; in contrast, the equivalent figure for the PIO, which obviously deals with a number of historical enquiries from the public and Members, was 56%.

After PARL83, EDM was the file next most heavily used. Overall it hosted 11% of the on-line sessions, though the figure climbed to 21% in the case of the PIO. Indeed, after the Table Office, the generators of the database, they were its biggest customers. There was one possible explanation for the large amount of use made of the EDM database by PIO staff. Apparently, Members issue press releases to local newspapers informing them of the EDMs they have tabled. The idea is to show their constituents that they were active in the House. As a result of these press releases appearing in the papers, people phoned up the House of Commons requesting more details, and were put through to the PIO. By contrast, Library Research staff fielded hardly any Early Day Motion questions. This could be simply a case of information switching; that is, enquiries are always directed to the specialist unit/section concerned.

A breakdown of file preference by unit discloses the following features:
1. Reference Room's marked preoccupation with CURRENT. CURRENT accounted for 94% of their use of all POLIS files and 96% of their use of the three POLIS parliamentary files. Given that the Reference Room's prime sources of information are non-parliamentary in nature, and given CURRENT's extensive coverage of this material, these statistics should come as no surprise. There are other reasons, too: staff coming to use a terminal expected it to be in CURRENT - and cursed anyone who left it in another database; and quick reference work invariably concerns current affairs most of the time - that is where MPs' interests lie.
2. The historical requirements of Statistics - more than 35% of their searching was conducted in the two archival POLIS files. This is attributed to their need to identify runs of statistics.

3. The heavy use made of the EDM file by Oriel Room staff (accounting for 15% of all use), who were obviously the principal recipients of Members' EDM enquiries (i.e. those made in person) and the Branch Library (15%), who no doubt take on the burden of the Research Assistants' EDM enquiries.

An analysis of file selection by the amount of time spent in each file, produced a very different result: CURRENT was still the principal file but become far more dominant. 97.5% of the time that Library users were on-line to POLIS, they were in CURRENT, so the sessions in the other files were of a much shorter duration. The average length of a session in CURRENT was 173 minutes, but only between 9 and 11 minutes in the others. Significant amounts of time were spent in two other files - 258 hours in PARL83 and 190 hours in EDM, but their use pales by comparison with the enormous amount of time spent connected to CURRENT (22,432 hours), which was obviously home base and the first file to be interrogated. It was also the biggest.

TABLE 4.12
POLIS FILE USE: CONNECT TIME[1]
(Logs)

		Current	Parl83	Parl79	EDM	HOLLIS
END-USERS	%	91.4	5.2	3.0	0.2	0.2
	ave.	65	19	20	13	26
Branch Library	%	91.7	5.0	3.1	-	0.2
	ave.	129	22	24	4	26
Dial-up	%	88.9	6.9	2.4	1.9	-
	ave.	10	10	6	13	-
LIBRARY STAFF	%	97.5	1.1	0.5	0.8	-
	ave.	173	11	10	10	9
Research	%	98.1	1.4	0.5	0.1	-
	ave.	171	15	13	12	3
Reference	%	97.7	0.8	0.5	1.0	-
	ave	201	8	11	11	12
PIO	%	95.6	1.3	0.8	2.4	-
	ave.	128	6	7	8	-

1. in minutes

Employing sessions conducted as the measure, at the broadest level end-users were more single file oriented than Library staff, but only just. Even so there were many individual Library user groups who were even more single minded - the Reference Room, for instance, and if we look just at the Branch Library end-users, they were one of the least single-file oriented of all the user groups. According to the other

measure - time spent in each file - Library staff were more single source oriented than the end-users. Dial-up users were again the odd ones out, being one of the most single source users when session numbers were considered and the least single source when time was being considered. End-users spent proportionally more of their time searching retrospective files than Library staff.

4.3.3.4 Types of search conducted

According to Branch Library staff members, who because of their proximity to the POLIS terminal, observed a lot of end-user searching, MPs' Research Assistants undertook three types of POLIS search: subject searches; what they called *global* searches (searches on physical form, with no subject specified); and searches on a Member's political opponent(s). The survey provided strong confirmation of the popularity of the first two search types, and somewhat less support for the popularity of the third. Straight subject searches (e.g. on topics such as defence, higher education funding, student loans, and poll tax) were popular, with 62% of the respondents saying that they conducted one of these searches frequently or very frequently (Table 4.13). Research Assistants would often sample POLIS to find out what was available on a subject, to get a feeling for its weight and size: throwing in terms like capital punishment or defence. Obviously, the type of information that was sought depended on the nature of the Member's field of work or interest. Thus there was the case of a Research Assistant who worked for a Member who was interested in the *Children's Bill*, and who had frequently to search for information on adoption. He has moved since to another MP, who has different interests, and no longer searched for anything to do with children. And Research Assistants do move around MPs of the same party.

TABLE 4.13
TYPE OF POLIS SEARCH CONDUCTED BY RESEARCH ASSISTANTS
(Survey)

	V. frequently		Frequently		Occasionally		Seldom		Never		Total	
Search type	n	%	n	%	n	%	n	%	n	%	n	%
Subject	13	23	22	39	15	27	5	9	1	2	56	100
Subject/MP	8	14	16	29	21	38	10	18	1	2	56	100
MP	1	2	3	6	21	39	28	52	1	2	54	100
Parliamentary Questions.	20	35	20	35	14	25	3	5	-		57	100
Progress of Legislation	3	5	7	13	22	39	21	38	3	5	56	100
Other Parliamentary proceedings	7	13	10	18	25	45	13	22	1	2	56	100
Committees & Working Parties	2	4	5	9	27	50	17	31	3	5	54	100
Library materials	3	5	7	13	21	38	17	31	7	13	55	100

Research Assistants also said that they used POLIS as a quick reference parliamentary tool, to answer queries like the date of a Member's speech on the Children's Bill, or how many questions have been asked on special education. This showed up strongly, for searches involving Parliamentary Question (PQs) were shown to be the most popular of all searches, with: 70% of respondents saying that they conducted them frequently or very frequently. Interestingly, as we shall learn later, the logs did not really bare this out, for they showed that subject searches were many times more popular than PQ searches. Of the least popular searches were those just on MPs names - an undoubted surprise, and Library Materials, not so surprising perhaps, given that some Assistants would not know what this term meant (the category did in fact include material that would be useful to Assistants, like Library briefing papers). The person who said they never searched on subject must be something of an enigma.

Turning to the computer logs now, the data obtained here on which fields were searched probably holds the most interest of all the transactional log data, especially given what has been said in the literature about the preference of the subject search and the limited search approaches taken by end-users. Through such data, we can learn not just the approaches users took to searching the POLIS database (and, through this, infer what kinds of questions they were asking), but, also, whether they preferred field-specified or default searching. Were for instance the big investments made by the Library in maintaining a controlled language database fully justified by user search behaviour; were both intermediary and end-user beneficiaries, and was the large and complex field structure of POLIS fully utilised?

In comparing the search approaches of the two main groups it has to be borne in mind that virtually all intermediary based searching was conducted in command mode, whereas end-users mainly, but not exclusively, used the menu-driven interface. Essentially then the end-users search options were partly prescribed and narrowed by the system. Table 4.14 shows the manner in which they were prescribed.

TABLE 4.14
POLIS MENUS

Welcome to the menu system, do you wish to :-

A	Perform a guided search (Menu system).
B	Call another database
EXIT	Search the database freely (experienced users only)
LOGOUT	Sign off the system

MAIN SEARCH MENU - MENU 1.1

Do you wish to search for:-

A.	Parliamentary Questions
B.	Progress of legislation (UK and EC)
C.	Other Parliamentary proceedings and papers
D.	Information about committees, working parties etc.
E.	Library material (including official publications, books and deposited papers)
F.	All the above
G.	Call another database
EXIT.	Leave the menu system
LOGOUT.	Sign off the system
BROWSE.	Look at the Thesaurus

Please enter selection > e

                                                              Working

SEARCH MENU - MENU1.2

LAST SEARCH - 44536 - group=(op or lib or c or int) or type=eur

Do you wish to narrow your search using the following:-

A.	Member (or Peer) asking/ Minister answering	EXIT.	Leave menu System
		LIST.	List your searches so far
B.	A Subject	LOGOUT.	Sign off the system
C.	A Government department, Corporate Author, etc.	BROWSE.	Look at the thesaurus
		X.	Start search again
D.	Date	Y.	Display current details
E.	Session	Z.	Return to previous menu

(MENU1.1)

F.	Name of Bill or Act
G.	Personal Author
H.	EC Document Number
I.	Names
K.	Titles

Please enter selection > g
Search the AU field

There were 1904 occasions when a field was used in searches - in the case of a default subject search a number of fields would be used. A wide range of different approaches were made to POLIS by the Research Assistants. Altogether 14 different fields figured in the searches of the Assistants. (Table 4.15). Either the restricting effect of the menu was at work here (just 15 choices were given in the menus) or end-users simply used fewer approaches, for House of Commons Library staff used 27 fields altogether. Branch Library end-users employed (three) more fields in their searches than their dial-up colleagues - illustrating yet again, that despite the greater opportunities they had to search POLIS in command mode, the latter group proved not to be the expert searches that some Library members had initially thought.

<div align="center">

TABLE 4.15
POLIS FIELDS SEARCHED
(Logs)

</div>

	Branch library		Dial-up		End-users		Library staff	
Fields/approaches	n	%	n	%	n	%	n	%
Default (IT)	1132	72.1	240	71.6	1372	72.1	4921	45.8
Publication (GROUP)	221	14.1	27	8.1	248	13.0	308	2.9
Subj. indexing terms (SIT)	79	5.0	31	9.3	110	5.8	1132	10.5
(TYPE) of publication	41	2.6	11	3.3	52	2.7	802	7.5
Members (MEM)	35	2.2	9	2.7	44	2.3	319	3.0
Organisation (ORG)	28	1.8	12	3.6	40	2.1	208	1.9
Date (DATE)	13	0.8	-	-	13	0.7	38	0.4
Session (SES)	7	0.4	-	-	7	0.4	37	0.4
Corporate author (COR)	3	0.2	2	0.6	5	0.3	333	3.1
Legislation (LEG)	5	0.3	-	-	5	0.3	159	1.5
Title (TI)	2	0.1	2	0.6	4	0.2	398	3.7
Others	3	0.2	1	0.3	4	0.2	2082	19.3
Total	1569	100	335	100	1904	100	10737	100

N.B. It was not possible to distinguish between single field searches and hybrid ones.

4.3.3.4.1 Subject searching

POLIS offered two types of subject search: there was the default subject search (i.e. one with no field qualification - in effect, a *free text* search) and the controlled one, using terms from POLIS's sizeable thesaurus. The default subject search (IT) - representing a search on six subject-related fields: description (DES); subject indexing terms (SIT) - words or phrases taken from the POLIS thesaurus; new indexing terms (NIT); organisations as subjects (ORG/ORG1); and names as subjects (NAME); was by far the most popular end-user approach, accounting for some 72% of the fields mentioned. This was surprising, for as has been mentioned POLIS' menus did nothing to encourage or help the subject searcher. Indeed, the first menu (Table 4.14) did not even mention that there was a subject approach. The main choices offered by the lead menu clearly showed the system's origins and prime concerns: the indexing and

location of parliamentary documents. If you required a subject search you would have to opt for choice F, which was vaguely labelled, *all the above*. It was by no means an obvious option for the Research Assistant with a subject search in mind, but according to the Library staff, regular users of the menu system soon caught on: *they're bright kids and hit the F key as a matter of course*. It was only at the second menu that a subject choice was offered - and even then it was buried amongst another 11 choices with, significantly, the Member search topping the list - something especially important in the light of the Member field analysis that comes later. Clearly, Assistants were not noticeably influenced by the menus here, they simply overrode them.

It would appear then, that the menus were not designed around the information-seeking preferences of researchers. POLIS staff admitted as much, and conceded that the second menu would perhaps have been a better first menu. The menus, of course, were designed with little data to hand (a system has to be constructed first before you can assess its effectiveness). Since the completion of this research the menus have been changed to take account of the popularity of the subject search. Interviews and observation with Library Staff had indicated that the typical way Research Assistants used POLIS was to throw in subject terms and to find out, through broad displays of data, what was of interest, and this would appear to be borne out by the monitoring data.

By comparison, Library staff used the default search fewer times - 46% of the times in all. Though the Research Assistant's figure was not far removed from that of the International Affairs section, where default searching accounted for 70% of field searches. So again we have a case where there were similarities in the searching behaviour of intermediaries and end-users.

The controlled language field (SIT) was the end-users third most popular field - the publication group search was, in fact, second. Research Assistants used this field around 6% of the times. While this hardly represents heavy use of the controlled approach, even so, this was more than a little surprising, for the SIT search was not an option provided by the menus - only a default subject search could be undertaken. That provides hard evidence that Research Assistants entered command-driven mode, where the SIT search was available, to conduct these searches. These Assistants probably woke up to the difficulties of conducting a precise subject search though the menus, and looked at the POLIS guide, which was placed next to the terminal, and realised how easy it was to do a subject search in command-driven mode (the POLIS

guide is really a guide to command-driven POLIS). Also each menu (Table 4.15) reminds the searcher that one option is to leave the menu-system - it is inviting of course, and anyone with an inquiring mind is likely to have a go (though admittedly, they were not told where they were going to when they left the menu system). The amount of controlled language searching conducted by Research Assistants was below that for Library staff, who employed the SIT field 10% of the times. There was not a lot in it, however, and, given the same interface or a better subject menu, there might have been no differences in it at all (see below).

Thus, dial-up users, who had a genuine choice as to which mode they searched POLIS in - their computers were not perpetually parked in menu-driven mode - showed a markedly stronger preference for the SIT search: use climbed to 9.3% in their case - a higher percentage than some of Library Research staff units, with whom they had most in common. Now, this could be a reflection of their generally superior on-line skills, which has been previously alluded to (though the data so far was not wholly convincing on this count), but more likely, it was because - unlike Branch Library users - they entered the system in command driven mode and it was sometimes quicker to enter, say, *Find SIT=child abuse* rather than */menu* to get into menu driven mode and then hit the F key and then wait to enter the subject terms. They have to know about the slash command and tramp through two menus, so why not use the SIT facility instead?

Generally then, the subject search was the much preferred one: used in 78% of the cases if IT and SIT searches were combined, and rising to 80% if we include ORG (as a subject) and NAME (as a subject) searches. Interestingly, when questioned about the amount of their subject searching (see page 110), Research Assistants understated their use of the subject search - less than two-thirds said they searched on subject frequently or very frequently. Perhaps, respondents assumed that they were being asked about a pure subject search, i.e. one without form or date qualifications? Research Assistants favoured the subject approach more than the House of Commons Library staff, who employed the IT and SIT fields only 56% of the times. However, despite the approach's popularity with end-users, it did not follow that the approach was an easy one for them: after all the survey pointed to one-third of the researchers regularly encountering difficulties in formulating the search enquiry (Table 4.21). Too often they found that they did not have the right balance - if they were too broad, they had to page through large numbers of documents and, if too narrow, they had to start the search all over again - not being able to return to the previous menu level. Controlled language indexing also came in for complaint: a number of users said that,

while the terms chosen to index a document were probably technically correct, they were not the ones they would use.

Comparison with Library staff.

With the command-driven system Library staff had a completely free-choice as to how they would conduct the subject search. Despite the size of the controlled language - approximately 9000 words, and the fact that many staff had once worked in the Indexing Unit (where staff obtained a close familiarity with the controlled language), the default search was, by far, the most favoured. Close to half (46%) of all searches were made in default mode. It could be argued, that in the case of the Library staff, the use of the default search was not as uncontrolled or loose as might be imagined, for being experienced users, they would tend to know from past searching experience what terms have been used to index documents and, anyway, the description field on POLIS - essentially the only natural language field - was in many respects standardised, using highly formal governmental and Parliamentary language.

Second most popular in use was the other subject field - the controlled language (SIT) field, which accounted for just over 10% of field searches. So the two most popular approaches to POLIS were subject searches, together forming over 56% of all searches. What is puzzling about this is that staff must have been aware of this when they designed the menu-interface for Research Assistants, so why did they think that Assistants searched differently from themselves?

TABLE 4.16
POLIS FIELDS SEARCHED: MAIN LIBRARY STAFF DIVISIONS (Logs)

	Research staff		Reference staff		PIO		TOTAL	
	n	%	n	%	n	%	n	%
Default (IT)	2559	54.4	1010	29.1	1352	52.7	4921	45.8
Subject (SIT)	430	9.1	459	13.2	243	9.5	1132	10.5
TYPE of publication	185	3.9	413	11.9	204	7.9	802	7.5
Description (DES)	302	6.4	337	9.7	66	2.6	705	6.6
Title (TI)	139	3.0	167	4.8	92	3.6	398	3.7
Corp. Author (COR)	120	2.6	154	4.4	59	2.3	333	3.1
Members (MEM)	32	0.7	203	5.9	84	3.3	319	3.0
Publication GROUP	83	1.8	157	4.5	68	2.6	308	2.9
EC Doc. No.	239	5.1	4	0.1	4	0.2	247	2.3
Reference (REF)	68	1.4	71	2.0	93	3.6	232	2.2
Author (AU)	118	2.5	81	2.3	10	0.4	209	1.9
Organisation (ORG)	76	1.6	63	1.8	69	2.7	208	1.9
Identifier (ID)	71	1.5	63	1.8	52	2.0	186	1.7
Legislation (LEG)	26	0.6	79	2.3	54	2.1	159	1.5
NAME	42	0.9	24	0.7	34	1.3	100	0.9
Others	210	4.5	185	5.3	83	3.2	478	4.5
TOTAL	4700	100	3470	100	2567	100	10737	100

In discussions with Library staff it seemed that the choice of which of the two subject approaches was taken was dependent on a number of factors, namely: (1) experience of the indexing process: the more familiarity with the indexing language the greater was its use - ex-indexers had, of course, the greatest familiarity and longest memories; (2) the nature of the enquiry; (3) familiarity with the subject of the query: when familiar it was easier to select the precise term - however, if unfamiliar, the pressure was on and the default search then offered more space and browsing room (and security); (4) the amount of detail provided: the more the detail the easier it was translate the query into an indexing term; (5) partly related to the latter, the vagueness of the query: controlled language searching was too precise an approach when the query - as it was so frequently - was couched in imprecise terms.

An examination of the figures for the principle Library divisions and individual units, shows there to be large and sharp differences in practice. Research staff (few indexers amongst them) were far more likely to use the subject default search than their Reference colleagues: 54% of their searches were conducted in the default mode, whereas the figure for Library Reference division (lots of indexers there) was very different - 29%. There were two main reasons why Library Research staff used the free text search more frequently: firstly, they, like MPs' Research Assistants, were far more ignorant of the controlled language - few had worked in the Indexing unit, and; because POLIS was not so central to their work, they searched it less frequently and consequently retained less knowledge of it. Secondly, as with the nature of research and advisory work, they did tend to get the vaguer queries - in consequence, preferring the width afforded by the default search.

Looking at the performance of the individual Library units (Table 4.17), International Affairs was the biggest user of the default search - 70% of their searches were of this type. In total contrast, Home Affairs, only occasionally used the default search - just 12% of the time. In the case of International Affairs their high use was mainly explained by their almost total rejection of controlled searching - they used it less than any other unit. Its lack of use had nothing to do with frequency of use, for International Affairs were one of POLIS's biggest users, it was simply a case that the Unit had a very strong free-text culture. The situation pertaining in Home Affairs was explained by the fact that, although they used controlled language searching the most of all the units, they were also simply not big subject searchers.

TABLE 4.17
POLIS FIELDS SEARCHED: LIBRARY UNITS [1]
(Logs)

	Econ.	Educ.	Home Affs.	Internat.	Science	Statistics	Oriel	Ref. Rm.	Branch
Default (IT)	42	61	12	70	21	41	23	19	51
Subject (SIT)	7	14	17	6	12	12	13	16	12
TYPE of pub.	4	3	3	4	3	3	14	10	7
Description (DES)	7	1	9	-	12	20	12	7	4
Title (TI)	12	2	16	-	2	3	5	12	2
Corp. Author (COR)	7	2	2	2	2	2	5	6	2
EC Doc. No.				10				2	
Publication GROUP.	1	2	1	2	7	3	5	1	5
Members (MEM)	-	2	1	-	4	1	6	7	4
Author (AU)	6	-	19	1	-	2	2	7	2
Reference (REF)	2	-	8	1	1	3	3	1	1
Organisation (ORG)	3	4	3	1	1	1	2	3	2
Identifier (ID)	3	4	6						
Legislation LEG)	2	1	1	-	3	-	3	2	1
NAME	1	3	3	-	1	1	1	2	1
DATE	1	1	1	-	1	1	1	1	-
STATS	1	-	-	-	-	4	-	1	-
NUMBER	-	-	-	-	24	-	1	1	-
Others	1	-	-	3	6	3	4	2	6
TOTAL	100	100	100	100	100	100	100	100	100

1. expressed as a percentage.

What of the controlled language approach? Well, in strictly numerical terms Oriel Room staff were the biggest controlled language searchers (332 searches), which befitted their status as the Library's most sophisticated and skilled POLIS user - many of their staff having worked in the Indexing Unit. International Affairs (133) and Statistics (127) were the other big users. As these were all high-volume users, it does make one wonder whether it is as much heavy use as professional training which determines whether the approach was used. In proportional terms, though, Home Affairs (despite a strong anti-POLIS/on-line stance of its Researchers) topped the list - no doubt the influence of their two Library assistants - former indexers - was at work here. It seemed that the units which used the controlled language search most frequently (e.g. Home Affairs, Reference Room) used the default search least frequently, which seemed logical.

Combining the two forms of subject search showed, that for two units - Education and International Affairs, around three quarters of their searches were of a subject type whereas for another Unit - Economics - subject searches represented less than half of all searches. The explanation for the unusual pattern of searching in the Economics Unit lay with their general unhappiness with POLIS.

The BASIS monitoring software also provided evidence of the individual thesaurus (SIT, NIT) words/terms used, but because the data was sampled (11 days throughout the survey year) and because of the inevitable scatter of use (there were more than 9000 words/terms that could have been used), there was insufficient data to provide a meaningful breakdown of the figures. Cumulating the data for all user groups showed that 1522 SIT/NIT terms were used on the 11 sample days, about 17-18% of the terms available. Despite the wide variety of terms used there was still a noticeable concentration of use, with just 23 terms (4% of those used) accounting for nearly a quarter (23.5%) of all those used. The terms/stems (*) that obtained the most use (more than twenty uses in the year) were: Europe*, security, communication*, child* and civil service: a list of common and fairly broad terms that contains no real surprises.

Library staff used a much wider variety of approaches than end-users - this was in part due to the restricting view of the menus and in part due to the bibliographic nature of much library searching. Following on from this, they were also not as preoccupied with the subject access as end-users were. Controlled language subject searching was not, however, just the province of the Library staff. Despite its absence from the menus, dial-up end-users in particular, used the facility as much as Library Research staff. It was though, the former indexers that revelled in its use - especially those in the Oriel room. There were big differences in take-up between Library staff, with some units using it proportionally twice as much as other units.

4.3.3.4.2 Non-subject searching

The survey found that, of the five types of information that can be selected by users from the POLIS menu (Table 4.14), Parliamentary Questions (for instance, on the sale of arms to Iran) were most frequently chosen. 70% of the users looked up this kind of information frequently or very frequently. Searches on Parliamentary proceedings and papers came next, with 30% of the users conducting this search frequently or very frequently. The other types of information on the leading menu - legislation, committee or working party reports and library materials were said to be used only occasionally. The low level of demand for non-parliamentary proceedings information may be due to the following factors: lack of knowledge on what is available within these categories of information (which may be related to lack of training) and the availability of other means of finding these types of information (e.g. Research Assistants could look up the *Weekly Information Bulletin* for information on Bills).

The computer logs confirmed that publication form searches were relatively popular with Research Assistants: publication type (TYPE) and publication group (GROUP) accounted, respectively, for 2.7% and 13% of all the fields used: 15.7% in all (Table 4.15). TYPE searches operated on individual documentary forms, like Library briefing papers, whereas GROUP searches were conducted on broad bands of documents, like Library materials, of which briefing papers, for instance, were just one type. It would be difficult to conceive of another on-line database in which form searches were made so frequently, but then physical form holds a great deal of significance when it comes to Parliamentary affairs and legislation - and especially so, it seems, for Research Assistants, because they employed the facility more than the Library staff, who might have been expected to be more form oriented. For the Library staff, the TYPE and GROUP fields accounted for only 10.4% of the fields used. Again the menus, which almost required the searcher to introduce form into the search statement, must have had some influence on end user searching - the search for Parliamentary Questions was first on the menu. The menu system, actually made a strong presumption that it was form that the researchers were principally concerned with.

Another surprising feature of the field analyses was how few times Research Assistants searched on Member. The survey results and interviews with librarians had suggested that researchers were very Member-centred (e.g. wanting to know everything their Member or opposition number said in a session, on a subject, during a debate etc.). About two-fifths (42%) of Research Assistants said they *sometimes* looked up a subject that was associated with an MP. POLIS treats all Members as authors (that is, authors of Parliamentary Questions) and it is obviously useful to check on an MP's contributions now and again. It was, however, comparatively rare for Research Assistants to search for all mentions of a given Member (dubbed an *incestuous search*): only 7% said they did this. In fact, the logs failed to provide confirmation of this, for the Member (MEM) field was represented just 2% of the times. It was an option missing from the first menu, though it does figure prominently in the second menu (it is top of the list of choices). Indeed, the Member search was not particularly common with other groups of POLIS users either, accounting for less than 1% of the fields used by Library Research staff, but rising to 6% in the case of Reference staff.

Other surprisingly lowly used fields, were corporate author (used 0.3% of the times) and date (0.7%). Neither of these fields could be considered as peripheral, date in particular is regarded by most information retrieval systems as a key access point:

perhaps, in the case of date, rather than try and wrestle with the precise form of input, it was better to make use of the reverse chronological display feature and browse your way through until you find the date required. Corporate author searching can prove difficult - there is a citation order to get right and stop words to consider. An approach used extensively by Library staff (6.6% of the times), but denied to the menu-driven searchers (and consequently not used), was the description field one. The subject default search does search this field though.

For Library staff the individual forms of publication (TYPE) field was the next most popular field - representing 8% of the fielded searches (Table 4.16). The preferred end-user form search was the GROUP one, so here was a case of Librarians seeking extra precision or demonstrating their greater knowledge of the Byzantine form classification. Given that Librarians were generally seeking documents and Library Researchers seeking (subject) information, we would expect this to show up in the relative use of the form/subject fields - and indeed this was largely the case. No Research Unit searched the form fields as frequently as the Library Reference Units. The form of publication search obviously had its place in the Oriel Room - 19% of the fields they used were of this type, and they accounted for 41% of all Library staff's use of these fields. By way of contrast, only 5% of the fields searched by Education were form fields. A stereotypical picture of a library professional searcher would portray them as controlled language and publication form searchers - this was indeed the case.

Following on in importance, were the description (6.6%), title (3.7%), corporate author (3.1%) and Member (3%) fields. In proportional terms, Statistics employed the description field search most - their favoured approach was to page through documents' description fields to obtain what they call the context of their statistical enquiries. The author approach was seldom used by House of Commons Library staff for, as well as that relatively low use of the corporate author field, the personal author field featured only 1.9% times. This is quite unusual, for after all, POLIS is a library catalogue and such catalogues are typically designed to cater for the author search: certainly a good deal of provision was made for such searches. Personal author searches did cause problems on POLIS though - you have to ensure that the form of citation is correct (e.g. punctuation, initials, and inversion all have to be considered). Though this fact could not explain its low-use amongst end-users, for the menus display the correct form of citation with an example. Home Affairs, though, were the exception for they obviously had a special fondness for the personal author field, using it 19% of the times. Science, by contrast, did not perform any personal author

searches. It was quite possible that, because author searching was so 'fiddly', users switched to a title search instead (the title search is reasonably well-supported - 3.7% searches overall). Though with author and title searches combined only accounting for 8.7% of all searches it did seem that few users brought any details to the terminal with them.

The relatively low number of searches conducted on Members' names - already noted in the case of Research Assistants - was mildly puzzling; one would certainly have expected them to account for more than 3% of field searches. Significantly though, the Oriel Room, the part of the Library closest to the Chamber, used the Member field most frequently - they accounted for nearly half (49%) of the times the field was used. Also, given the nature of much of the parliamentary business, it might have been expected that the legislation field would have figured more significantly - but it was only used on 1.5% of the occasions.

The absence of any significant date searching amongst Library staff as well as end-users, was also unexpected. On POLIS, date/chronological searching can be accomplished on three fields - session (parliamentary), month and date. If searches on all these three fields were combined, this still represented less than 1% of all searches. There were three methods of searching for date and untold ways of specifying date, so, perhaps, users were not confident about this search, especially if they had to do it quickly. Date searching could also be very slow and speed of delivery is everything when you have very important people standing over you waiting to appear in the Chamber or on television.

POLIS made quite extensive provision for searching statistics. Three types of statistics can be searched for - national, local and long-runs. As we have heard, Members frequently require statistics as ammunition to use against the Opposition party. However, such searches hardly feature at all in the analysis. Apparently, Library staff were discouraged from dealing with statistics as such questions are largely left to the Statistics section, and it can be seen from the Table that they do use that facility (but really not sufficiently to justify its provision). Finally, the Oriel Room staff, who were mainly professional librarians, employed more fields in their searches than anyone else. In their case 11 access points attracted more than 50 uses. International Affairs, with 4 fields attracting over 50 uses, came closest to them. Not unexpectedly, Education and Home Affairs, who are the two sections least enamoured with POLIS, employed the fewest fields.

The most highly trained users at The Commons made most use of the non-subject fields. The publication GROUP search was the only non-subject field searched proportionally more by end-users, but this was not surprising as that their opening menu consisted largely of publication group options. This might have explained the absence of author and title searching too. The computer logs appear to explode the myth of the popularity of the Member search. Of the Library searching groups what was most noticeable was the fact that they searched very different fields: there was no common search pattern - making generalising about intermediaries not just difficult, but also dangerous.

4.3.3.5 Use of system commands

A search of an on-line system usually involves much more than entering a term or terms together with an appropriate search command (**find** in the case of POLIS). It is also generally possible to truncate your search terms (the asterisk has this function on POLIS), range (usually performed on date and accomplished on POLIS by using the colon), or inspect an alphabetic display of items for suitable terminology (the **look** command). POLIS offered one additional, relatively unique feature, the switched search. This was used to switch searches from a non-preferred form to the preferred form. This was done automatically and the user was generally not aware that the switch had occurred. The switching facility was used mainly in the case of subjects (for synonyms), organisations (for abbreviated forms of an organisation's name) and Members' names (a surname search will be switched to the full form of the Members' name). The BASIS monitoring software tracked the use of all these search features, though, as was the case with fields, the data was only available for the busiest day of the last 10 working days of each month.

Of the searches conducted by Research Assistants, the vast majority (96%) were simple, straightforward searches, involving none of the above search facilities: the figure would be even higher (99%) if the switched searches were excluded, for the user would have had no control over this and would have known little about it. The simple searching might well be a reflection of a lack of search sophistication but it could be a product of searching a menu-driven system. It was not possible, for instance, to inspect the alphabetic inverted indexes in menu-driven POLIS. The **look** facility was obviously invaluable in term selection and as such would be particularly useful to end-users, given the inter-disciplinary nature of the POLIS file and the softness of the current affairs vocabulary, but, there were technical problems with providing the facility with the menu interface. It was possible to examine the

thesaurus, but, unfortunately, the use of this facility was not monitored by the BASIS software.

By comparison, the Library staff, with their command-driven facility, did use the **look** command but interestingly only the Research staff used it significantly: on average, they inspected the alphabetic index every 25th search. This does partly demolish the argument put to the researcher by professional Library staff that Research staff's preferred style of searching was to dive straight in by 'shooting' in a term and then browsing around: no looking up thesaural or alphabetic lists of terms for them. According to one librarian: *they just want to go to it, type in the term and get the answer - it is in their personality to do so.* Possibly, the Reference staff, many of whom were professional librarians who had worked in the Indexing Unit, were much more familiar with the vocabulary and, consequently, were less likely to use it. This does suggest the facility would have been useful to the end-users.

With high levels of free-text (IT) searching on show, one might have expected heavily truncated searching too, for both types of searching help overcome a lack of knowledge of the system (its vocabulary and field structure). However, this appears not to be the case. Research Assistants used the truncation facility less than 1% of the time. The very low use of truncation (used much more by Library staff) must surely have had its impact on the effectiveness of searching, especially when given the high levels of subject default searching undertaken by Research Assistants. There was little difference between the two groups of Assistants. By comparison - and speaking proportionately - Public Information Office staff used the facility the most (4.6%). Perhaps the explanation for truncation's poor showing was due to the fact that the truncated search was not as easy to conduct as it seemed: it does require a knowledge of what truncation involves and what symbol should be used to perform the function (and the explanation is buried in the POLIS Quick Reference Guide: menus provide no note or reminder of it).

TABLE 4.18
POLIS SEARCH COMMANDS USED (Logs)

	Standard searches		Searches with truncation		Searches that are switched		Searches with ranging		Displays of the index	
	n	%	n	%	n	%	n	%	n	%
END USERS	1834	96	10	<1	47	2	14	1	-	-
Branch	1514	96	8	<1	34	2	13	1	-	-
Dial-up	320	96	2	1	13	3	-	-	-	-
LIBRARY STAFF	9329	89	446	4	455	4	60	1	242	2
RESEARCH	4197	89	183	4	156	3	13	-	151	3
Economics	354	83	23	5	11	3	4	1	33	8
Education	299	90	17	5	6	2	3	1	8	2
Home Affairs	147	75	26	13	9	5	1	1	13	7
Internat. Affairs	2097	92	9	-	100	4	1	-	85	4
Science	164	88	13	7	8	4	1	1	-	
Statistics	940	91	73	7	14	1	2	-	4	1
C Room	199	84	22	9	8	3	1	-	6	3
REFERENCE	3010	87	146	4	202	6	26	1	86	3
Oriel Room	2121	85	113	5	161	6	24	1	74	3
Reference Rm.	158	83	11	6	9	5	1	1	11	6
Branch Library	731	93	22	3	32	4	1	-	1	-
Public Info. Off.	2322	90	117	5	97	4	21	-1	10	<1

Comparison with Library staff.

Nearly 89% of all interactions were what might be called 'standard' searches. An additional 4% were standard searches that were switched. Truncated searches, which might be used as a matter of course on a natural language system, played only a minor role in the searches of an essentially controlled language POLIS - just 1 in 24 included a truncated term. Ranging was little used (0.6% of searches). Finally, on 2% of the occasions POLIS Library users inspected the indexes. This relatively low figure could perhaps be explained by two factors: (1) users were always in a hurry (and did not have the time); and, (2) the user group was a well-trained and highly practised one, quite familiar with the vocabulary of the database.

The principal user groups searched differently. PIO staff, possibly through lack of time, hardly ever inspected the inverted indexes - they did so on just 10 occasions, whereas Library Research staff, probably the group least familiar with POLIS' vocabulary, were the biggest users of the facility, using it 151 times. However, the PIO obviously preferred truncation: proportionally, they used truncation most - nearly

5% of the times. Reference staff had the biggest proportion of their searches switched (nearly 6%) - no doubt a function of the relatively large amount of searching they did on Members' names (Table 4.16).

Relatively speaking, of the Library units the Branch Library and International Affairs were the biggest proponents of 'no frills' searching - 93% and 92% respectively of their searching was of the standard type, figures which rose to 97% and 96% when switched searches were included. By contrast, 75% of the Home Affairs' searches were of the standard type. Truncation was used heavily (13% of the time) by Home Affairs, which counted amongst its staff two former indexers (obviously indexers who were wary of the indexing system). They also made frequent examinations of the index (7% of interactions), though Economics (8%) just piped them for the position of number one index displayer. Not only were Home Affairs low users of POLIS, they were also cautious users - one probably follows from the other. Numerically speaking, Oriel Room used truncation (113 times), term switching (161 times) and ranging (24 times) most. International Affairs took first place for index displays (used 85 times). This could be explained by the indexing use to which their terminals was sometimes put.

As mentioned earlier the monitoring software did not enable examinations of the thesaurus to be pinned down to individual password holders, though it did provide global data on its use. From this data it was possible to distinguish between the use of the thesaurus through the menus and through the command-mode, so the evidence could be provided in a roundabout way. On the (strong) assumption that only end-users used the menus, Research Assistants used the thesaurus module 154 times in the year and Library staff 3874 times - twenty-four times as much.

The biggest differences of all emerge in this particular analysis of searching behaviour. End-users do use a narrower range of commands than Library staff, although this was partly related to the limiting affect of the menu and lack of training. There were no fundamental differences between end-users groups - and the fact that there were none suggests that the greater propensity of the dial-up users to use the command-driven POLIS makes no real difference to the range of commands used. There were, however, significant differences between Library groups. Thus on the one hand you had one group using truncation almost one in ten times (C Room Researchers) and on the other a group not using it at all (International Affairs). Library staff used the on-line theasurus significantly more times than Research Assistants

4.3.4 User satisfaction and problems encountered

When asked, end-users said that they were generally pleased with the results of their searches: 84% of them said they were usually satisfied with POLIS searches, while another 14% were satisfied with them some of the time (Table 4.19). Incredibly, only one of the 57 respondents, suggested that they were disappointed with the results that they obtained. There cannot be much higher praise than that.

TABLE 4.19
SATISFACTION WITH POLIS SEARCHES:
RESEARCH ASSISTANTS (Survey)

Usually	Sometimes	Not normally
48 (84%)	8 (14%)	1 (2%)

As a whole, users were pleased with POLIS. One respondent said, *they* (the results) *were beautiful*, and another commented, *I always find what I need*. Puzzlingly, all this did not accord with the perceptions of the Branch Library staff who, as a result of having the terminal located in their library, observed much of the in-house end-user searching. They were universally of the opinion that much of the on-line searching conducted by Research Assistants was unsatisfactory. To them - and the monitoring statistics supported them to only to some degree - researchers appeared to encounter all sorts of problems: typically, they came up with either nothing or vast numbers of items (something which was not supported by the logs), which they do not want to look through. The marked difference in perceptions could arise because Library staff had American researchers in mind - the least proficient end-users - when making these comments, and the sample survey only picked up a few American students. (The survey was conducted in July, when many had returned home). Possibly, too, Library staff had higher retrieval standards. Research assistants had lower expectations and as a result were more easily satisfied.

Despite the relatively high levels of satisfaction, over half of the respondents (58%) did experience occasional problems when using POLIS, although very few (9%), frequently experienced problems (Table 4.20). The fact that a third of respondents experienced no difficulties whatsoever was certainly significant, given that POLIS was far from being a simple or easy system to search, and that really nobody had been trained.

TABLE 4.20
EXTENT TO WHICH RESEARCH ASSISTANTS EXPERIENCED
PROBLEMS WITH POLIS (Survey)

Frequently	Occasionally	None
5(9%)	33 (58%)	19 (33%)

A problem common to many POLIS users was, unsurprisingly, that of access - 42% of respondents said so (Table 4.21). Many Research Assistants found that they had to queue during busy periods. A few voiced their criticisms of American students, who were accused of monopolising the terminals: *they spend so much time on the terminals as they are not sure of terms (e.g. What's a Session?) or do not know how to use the terminals*. What researchers most wanted to see were *more terminals* installed in the House. Some wanted more terminals in the Main Building or in the Branch Library; others wanted them to be in Member's offices, *on landings in buildings*, or in the Research Assistants' Room - in other words everywhere and anywhere.

TABLE 4.21
NATURE OF PROBLEM EXPERIENCED BY RESEARCH
ASSISTANTS SEARCHING POLIS (Survey)

Problem	no of problems	% of respondents (n=57)
Poor access	24	42
Equipment failure	4	7
The menus	16	28
Formulating a search	20	35
Printing	4	7
Response time	6	11
Total	74	100

One Research Assistant suggested that the terminals should be placed *in a quieter environment*, and another said, it was very *off- putting* to have somebody *looking over my shoulder while doing a search* (even the relatively proficient Reference Room Library staff did not like that). The vast majority of the respondents (96%) would like to have had office-based on-line facilities; one wanting it *now, if possible*. Given, the distance most had to travel to the solitary terminal, and the queues they frequently encountered when they got there, this finding was not very surprising. A few of them did realise that it might be difficult to have office-based facilities, because of the lack of space at the House - *I don't even have a place to put my things* and the cost - *I don't think it's going to happen in the near future because of the heavy cost involved.* Access problems were also mentioned in two other contexts. Firstly, there was the time available for researchers to conduct their searches. In an obvious reference to the

10 minute Library rule, one Research Assistant said that the *time* [he] *spent on a database now depends on restrictive policy*. Secondly, there was the time when POLIS was available to be searched. One researcher, possibly not aware of the dial-up facilities, wanted access to the terminals at all times - *even when the Library is closed*.

Once they had gained access, Research Assistants were confronted by another big problem - formulating their searches, 35% experienced problems here. The problem must be general and not associated with a particular host because virtually the same proportion experienced this problem with TEXTLINE (section 4.4.4). POLIS users often found their search enquiry was either too broad or too narrow - it was difficult to get the right balance. If the search term (e.g. *income tax*) was too broad, they had to go through hundreds of Parliamentary Questions, which proved *very frustrating*. They could narrow the search enquiry (e.g. from *education* to *special needs* to *between 1988/1989*), but they had to be careful not to make the search too narrow. If that happened, they could not return to the previous level, but had to go back to square one.

Research Assistants did, however, feel that the indexing of POLIS was part of the problem. They complained a lot about the way documents were categorised and indexed. They felt that, though the terms chosen might be technically correct, they were not those in common use (e.g: *electricity supply industry* instead of *electricity*), and therefore unhelpful to the user. They thought there was a need for better indexing and more cross-referencing: *Indexers do not think carefully enough on how users approach information* and *they should extend the range of vocabulary*. Some researchers even questioned the quality of indexing claiming that they could not find what they wanted, even though they knew the material was there. Given the highly controlled nature of the POLIS database and the limited opportunities to search on the natural language, it was probably inevitable that end-users become involved in controlled language issues: more indexing terms and better cross- referencing was a frequent cry.

As indicated earlier the POLIS menus were also a problem. While some researchers had no problems here: *It's easy to follow the menu*, a substantial minority did (28%): *they need rethinking: it's unclear, confusing; it's difficult to identify which type of parliamentary information one should access for a particular query*. For some, the options were too rigid: one respondent complained that he could only select one option or all options but not his own choice of options. As a result of their difficulties

with the menus, researchers felt that they *missed things* (which they knew were there) and had to ask Library staff to use the command-driven mode to get what they wanted. Respondents suggested that improvements be made to the system. They wanted a more flexible menu, that is, one with less rigid options.

In interviews with them, Library staff claimed that the slowness of searching the menus was also a source of dissatisfaction for the more experienced researchers. Indeed, two users did comment on the length of time taken to conduct a search using the menu system, one finding the response time *rather irritating.* and the other complaining that *it takes as long as 4 seconds to go into Option C* of the POLIS menu system. Seemingly, novice users were as time-sensitive as professional searchers and maybe more so, because they were not interested in the mechanics of retrieval and were generally less patient. According to one Librarian *everybody is in a mad hurry in this place.*

A few respondents wanted a fuller service from POLIS; for instance, they wanted the full-text of the answers to Parliamentary Questions put on the database - one of the most useful parts of Hansard. One of them commented that it was *very frustrating to know only the questions,* while another said that it was *not helpful just to have* [simply] *references to Parliamentary Questions.* Concise reports from committees were also asked for. All these researchers were asking for the full-text or abstracts - something Library staff have thought unnecessary or uneconomical given their close proximity of the hard copy. But Research Assistants did not have the hard copy collection close to hand, and as with all end-users appreciate the attractions of a one-stop service. Since the completion of this research the full-text CD-ROM of Hansard has been made available to researchers.

Users also experienced technical problems, like: equipment failure and slow and inadequate printing facilities (only a screenful of data could be printed at a time, and the limited number of printers available).

4.3.5 Training
Over two-thirds of the Research Assistants questioned learnt about POLIS from the Library tour, which most of them undertook on first joining the House of Commons. During the Library tour, the Assistants were given a very basic introduction to POLIS. 'Training' simply involved a brief demonstration on how to operate the POLIS menu-driven system. Users were not really told what the system could do for

them and were left to find that out for themselves. The whole session lasted for only 5 minutes. Not surprisingly, some Assistants were dissatisfied with the brief training they obtained. One commented that: *not much publicity* [was] *given to what the system can do,* while another said, that he had *not found out its* [the system's] *full potential yet* and, though therefore, that he *might take up Library staff's time unnecessarily.* Only 26% of researchers considered that they had any training at all in the use of POLIS. To the general surprise of the Library staff, who thought that perhaps they had misunderstood the question, six researchers said that they had been trained in the use of the command-driven version. Two of them belonged to the original group of Research Assistants who were given training after responding to a notice put up by the Library staff when POLIS was first introduced. It was not so clear how the others had been trained, but three of them had access to POLIS in their Member's office. The command-driven POLIS training session lasted for half a day and was described by those who attended as *very useful* and *helpful.* However, it was generally felt that refresher courses were needed; in the words of one of them: *It's the hands-on experience that counts; otherwise you'll forget the commands.*

One-fifth of the researchers learned about databases from their fellow researchers and secretaries, but significantly, only 7% heard about databases through their Members. One Researcher mentioned that he only discovered the facilities by chance. In addition to the very limited training opportunities provided, POLIS was also poorly publicised. A Librarian explained that this was Library policy and was necessary because of the lack of staff and terminals (Echoes here of *Guardian* management's ambivalent attitude to end-use - see page 162). The only form of publicity and advice on offer, therefore, were the leaflets on *POLIS* and *Information Technology*, which were distributed to all new Members and displayed in the Members' Library, and a POLIS 'quick reference' guide, located by the Library terminal, which frequently went missing. One Research Assistant suggested that the POLIS guide *should be chained to the terminal*, for without the guide, she did not know even the simple commands. The menus were not enough in themselves. Besides obtaining a full and proper training, Assistants wanted *a clear written manual, instruction sheets or handbooks*. One wonders why all interested researchers were not given their own POLIS guides? No doubt, resources would not allow it.

You cannot escape the feeling that, whilst searching by Research Assistants was not exactly discouraged, it was not exactly promoted either - a necessary evil? End-users were really left to sink or swim - and in this regard the number swimming must be regarded as something of a surprise. A small number of researchers expressed their

dissatisfaction with the state of things: *there is a lack of information on availability,* one bemoaned, and another simply wanted *published dates of [the] Library tour.* Given the lack of on-line support for users, it was perhaps surprising that such high volumes of searching were achieved and dissatisfaction so low. Perhaps, as one Library staff member suggested, it was because researchers were very *streetwise,* and quite capable of finding their way around. Compared to the extensive training given to the Library staff and the high priority it was accorded by Library management, the training given to Research Assistants appeared to be wholly inadequate. There was very little point bemoaning end-user skills - as some Librarians were wont to do - when they were provided with so little on-line support - what else could be expected?

Again, anticipating the *Guardian* journalists responses to come, but not voiced quite so strongly, almost one-third of the POLIS users expressed an interest in being trained. They felt that training would increase *the speed of* [their] *searches* and improve their effectiveness in using the system. Because of the lack of training, one Assistant said that he knew only *the basic methods of obtaining information,* while another said that she had to *stick to fairly straightforward searches* colourfully adding, *without training it's a bit like wallowing in rice pudding.* Another Researcher wanted to do his own searching but said *it'd be helpful if* [he] *were properly trained.*

In answer to the question, whether they would like to be trained in the use of the command-driven version of POLIS, nearly two-thirds of the Research Assistants (63%) expressed an interest. That was higher than the percentage who expressed an interest in being trained in the menu-driven version, so it has to be concluded that most researchers felt that they could manage the menus, even though they were not always that helpful. Those that did not require training were either temporary staff or students on short term placements. However, among the more established researchers there still was a worry that it would prove difficult *to take time from working to learn the system and to become fully familiarised with it.* Other reasons cited by those not wanting to be trained were the distance of their work-place (*I now work in Liverpool*) and, for a few of them, the adequacy of the menu-driven version (*It's not worth my while as I can use the menu system to do my searches*).

End-users at The House were really offered nothing in the way of training or information regarding POLIS. They managed by asking and watching their colleagues. Not put off by this, a large proportion of them wanted to be trained to search the system in command mode. For many end-users, but not apparently many librarians, menu systems are not an end in themselves: instead they were seen as a

bridge, or a leg-up, to more complex systems.

4.4 USE OF TEXTLINE

4.4.1 TEXTLINE: the information system

TEXTLINE, owned by Reuters, is a highly successful business information service, largely targeted at the City user. Its menu-driven software has particularly endeared itself to the business end-user. Its principal attraction for many subscribers at the time-was coverage of the British and international business press. TEXTLINE counts industrial, company and financial information amongst its strengths. While, once a wholly abstract-based service - albeit containing highly informative abstracts of a kind that would be difficult to distinguish from the full-text, over the years TEXTLINE's coverage has become increasingly full-text. TEXTLINE was, however, a highly selective service taking, largely business stories from UK national and regional newspapers, something, in fact, not all its users at the House were aware of. FT PROFILE would have been a better selection for current affairs devotees, like MPs and their Assistants, whose principal information resource was the newspaper, but the cost of running PROFILE as an end-user service was regarded as being too costly. There was also some evidence to suggest that Library staff wanted to keep PROFILE to themselves - it was such a useful and easy to use service. Indeed, if a recent House of Commons Information Committee (1994) report is anything to go by they are looking to exercise a monopoly over its use.

The House of Commons first subscribed to TEXTLINE in 1984. Its UK orientation, newspaper-based coverage and the relatively cheap unlimited use agreement on offer made it an attractive on-line option for Library staff. Its menu-driven searching facility meant also that it would be particularly suitable for direct use by Members' Research Assistants and it was introduced to Researchers in 1986. It is possible to search TEXTLINE in command driven mode, but it is really not set-up for it: command mode really offers a method of by-passing the menus and very frequent use would be required to search it in this mode. Chevrons, codes and a host of subject search approaches further complicates searching - see Table 4.22. TEXTLINE was accessed through two dedicated terminals, one located in the House of Commons Library Reference Room, for the exclusive use of Library staff, though it was alleged that a few Members have been known to search it themselves; and the other, which was intended for the use of Members' Research Assistants, located in the Branch Library, sited next to the POLIS terminal.

TABLE 4.22
TEXTLINE: SAMPLE SEARCH SCREENS AND MENUS

```
FULL SEARCH FACILITY
Search terms up to a maximum of TWELVE may be selected from:
* Company codes with optional "X" suffix
* Industry codes ("1" prefix) and Product codes ("PD" prefix)
* Topic codes of 1, 2 or 3 digits, with "T" prefix
* Country and Regional codes with optional "Z" suffix
* Source codes
* Key words or phrases
* Free text - single words or phrases
* Personal names - Surname:Initial:Title e.g. SMITH:A:SIR

Operating procedures
* Search terms must be separated by either a plus sign (+) for "AND", a comma (<) for "OR", or a minus (-) for "NOT"
* For increased precision and chevrons and code numbers as follows:-
Company code: <1    Ind/Prod code: <2    Topic code:        <3
Country code:  <4    Source code:     <5    Key word/phrase: <6
```

'General' databases	Updated to:		Updated to:
1. UK and W. Europe	03-2-88	6. Central and S. America	16.1.88
2. USA	28-1-88	7. USSR and E. Europe	31-1-88
3. Middle East	25-1-88	8. Africa	28-1-88
4. Far East	16-1-88	9. Canada	21-1-88
5. Australia and NZ	25-1-88	10 India	02-2-88
'Specialist' databases			Updated to:
A. Banking and finance	09-11-87	G. Engineering	31-12-87
B. Insurance and investment	24-12-87	H. Electronics and computing	24-12-87
C. Property	17-12-87	I. Accountancy and Tax	10-12-87
D. Marketing and media	23-12-87	J. Construction	11-12-87
E. Retailing	22-12-87	K. Travel	18-12-87
F. Chemicals	24-12-87	L. Aerospace and defence	30-11-87

```
Are the references to relate to material published:

1.        in the last month?
2.        in the lst 3 months?
3.        in the last 6 months?
4.        in the last 12 months?
5.        since a specified date?
6.        before a specified date?
7.        between two specified dates?
8.        since start of database?

Enter code number as required
```

TEXTLINE offered a genuine head-to-head comparison between Librarians and end-users: both groups could freely use the system, without cost considerations; and, unlike the case with POLIS, both groups largely used the menu-driven interface. The playing field was in fact even more level than that, for with POLIS, the Library staff had a distinct advantage in that they were never more than a few feet from a POLIS terminal, whereas Research Assistants invariably had a long walk to the terminal. But in the case of TEXTLINE both groups had a long walk. The Librarians did have one distinct advantage though, for they were the only ones that had received any training in its use. TEXTLINE also offered the opportunity to study the end-use of a text based system, a system inevitably more user-friendly - one that would provide more

than documentary signposts. While the logs were not as full as those of POLIS, they were probably more accurate, for TEXTLINE was a dial-up commercial service, with all the disciplines that this entailed.

4.4.2 The TEXTLINE users

The potential user group was as that for POLIS - some 600 Research Assistants and (possibly) a handful of MPs. The Library staff user group was though, much smaller, with just half a dozen staff trained to use TEXTLINE and many of them a good distance from the dedicated terminal.

4.4.3 Search characteristics
4.4.3.1 Use data
4.4.3.1.1 Frequency of use

The user survey identified 36 Assistants as having used TEXTLINE - 24 said they did not search the system. The population of actual users was then less than that of POLIS. At the very least then six times as many end-users searched TEXTLINE as Library staff. Use patterns for TEXTLINE were variable (Table 4.23), though generally TEXTLINE was not used as frequently as POLIS. This was not surprising for POLIS was much more central to researchers' every-day concerns (they said so see **page** x). After all, TEXTLINE was a business database. In addition, there was only one TEXTLINE terminal compared to the twenty two POLIS entry points. As many people used TEXTLINE regularly (more than once a week) as used it irregularly. Only one person used it daily, compared to the 6 people, who used POLIS daily. With high levels of irregular and infrequent use, and a total absence of training it was probably just as well TEXTLINE was menu-driven.

Branch Library staff, however, saw a markedly different pattern of use and said that TEXTLINE was in almost constant use all day long - a fact borne out by the monitoring statistics - which suggests that, if the survey was to be believed, a lot of people used it irregularly and were thus not picked up by the survey. According to Library staff as the Library opened, there was someone at the terminal. And it was not long before there was a queue: indeed, between 10 a.m. and 4 p.m., there was almost inevitably a queue, and it could be six deep. And the logs do show that on some days more than thirty searches were conducted, creating something of a bottleneck.

TABLE 4.23
FREQUENCY OF USE OF TEXTLINE BY
RESEARCH ASSISTANTS (Survey)

	n	%
Daily	1	3
More than once a week	9	25
Weekly	3	8
Fortnightly	7	19
Monthly	3	8
Irregularly	9	25
Occasionally	4	11
Total	36	100

The age profile of TEXTLINE users suggested that it was the younger researchers who used it (Table 4.24). 41% of the research assistants in the 18 to 30 age group used it at least once a week; the proportion dropped to 17% for those between 31 to 40; and nobody in the oldest age group used TEXTLINE that frequently - just as was the case with POLIS.

TABLE 4.24
FREQUENCY OF USE OF TEXTLINE BY RESEARCH
ASSISTANTS: ANALYSIS BY AGE (Survey)

	18-30	31-40	40+
Daily	1	-	-
More than once a week	8	1	-
Weekly	3	-	-
Fortnightly	6	1	-
Monthly	1	2	-
Irregularly	7	1	1
Occasionally	3	1	-
Total	29	6	1

4.4.3.1.2 Number of Sessions conducted

Research Assistants undertook a total number of 2340 TEXTLINE sessions during the year (Table 4.25). This meant that over the year they conducted many more searches on TEXTLINE than they did on POLIS - more than a thousand more. Of course the greater authenticity of the TEXTLINE logs could have had something to do with it. The highest number of sessions (265) were conducted, not surprisingly, in February, which was firmly established as the peak month for on-line activity (POLIS session data also peaked that month). In both June and July more than 250 searches were conducted, April (243) and May (242) were not far behind. A low of 97 searches was recorded in August.

TABLE 4.25
NUMBER OF SESSIONS CONDUCTED ON TEXTLINE:
MONTHLY ANALYSIS (Logs)

Month	Research Assistants		Library staff	
	Number of searches	% rise and fall on previous month	Number of searches	% rise and fall on previous month
Oct	167	-	90	-
Nov	158	-5%	54	-40%
Dec	112	-29%	40	-26%
Jan	166	+48%	53	+33%
Feb	265	+60%	70	+32%
Mar	214	-19%	88	+26%
Apr	243	+14%	n.a	n.a
May	242	<-1%	n.a	n.a
Jun	255	+5%	n.a	n.a
Jul	255	0%	n.a	n.a
Aug	97	-62%	n.a	n.a
Sep	166	+71%	n.a	n.a
TOTAL	2340	-	395	-

The overall picture was one of rapidly rising demand from the start of the Parliamentary session in November to the Spring, when demand reached its peak. Dips naturally occurred during the Christmas and Easter recesses. Demand held up fairly strongly in the early summer but plunged dramatically once the summer recess had begun. TEXTLINE use - as did POLIS use - would then seem to parallel the rises and falls of the House's workload.

Because of a technical fault on the Library staff's TEXTLINE terminal, data was only available for six months. During the equivalent six months 395 searches were conducted, just 17% of the Research Assistant total. On this measure the Library staff were comprehensively out performed.

Logically, it might have been expected that the Economics Unit would turn out to be TEXTLINE'S biggest user: the Unit's brief and TEXTILE'S coverage clearly matched. But while Economics did use the host daily, it was not used extensively - though it was thought useful for UK regional material (of all the hosts, only TEXTLINE covered the regional press at the time). Amongst members of the Unit, there was some disagreement as to its worth. With the TEXTLINE terminal sited in the Reference Room and knowing how the use of on-line databases, like any other information resource, can be boosted by physical proximity - it might have been expected that Reference Room library staff would be fairly heavy users of TEXTLINE. But, they were not happy with it and it was quickly becoming eclipsed by PROFILE.

Ironically, given what was said earlier, Reference Division staff actually thought that their low and declining use of TEXTLINE was partly due to its inaccessibility! While they called up PROFILE from their POLIS terminal, they had to change terminals (and sometimes move) to get access to TEXTLINE - proximity of course is relative. Surprisingly too, given the undoubted international strength of TEXTLINE, International Affairs, a big user of both PROFILE and POLIS, did not use it themselves. If they had a query they thought it could help with, they asked staff in the Reference Room to deal with it on their behalf, preferring not to search the system themselves. (Another instance of information switching at work).

Table 4.26 shows the actual number of searches conducted per day. The figures were based upon a year in which there was 250 working days. On one particular May day, 34 searches were recorded by Assistants. Perhaps, more telling, was the fact that on only 4 working days in the year was the TEXTLINE terminal inactive. The number of searches conducted varied greatly from day to day. The number of days which saw 1-5, 6-10 and 11-20 searches conducted was similar. Another testimonial to how busy that terminal got, was the fact that for 95 days (38%) of the year it was used more than 10 times.

TABLE 4.26
NUMBER OF SEARCHES CONDUCTED PER DAY
ON TEXTLINE (Logs)

Number of searches per day	Researchers. No. of days		Librarians. No. of days[1]	
	n	%	n	%
0	4	2	13	11
1-5	70	28	90	74
6-10	81	32	17	14
11-20	86	34	2	2
20+	9	4	0	0

1. Six months of data only.

By comparison Library staff usually searched TEXTLINE between one and five times a day. There were many more days (13) when the terminal was inactive, and many fewer days when fewer than ten searches were made.

4.4.3.1.3 Connect time

The logs tell us that Research Assistants spent 574 hours on-line to TEXTLINE in the year surveyed. (Table 4.27). TEXTLINE received about half the level of use of POLIS during the equivalent period, but given, that the method by which POLIS statistics were collected inflated its use, in all reality, the difference was probably nowhere near as large. To put the scale of end-use into true perspective, consider this

statistic, Research Assistant searching of TEXTLINE accounted for 55% of all external on-line searching undertaken at the Commons, by both Library staff and Research Assistants.

Library staff spent over 67 hours, about 11 hours per month, during the six month survey period. If these figures are extrapolated to the full year some 133 hours were spent on-line to TEXTLINE, a figure representing 28% of all external on-line searching undertaken by Library staff during that time. TEXTLINE was only second to PROFILE in regard to the time spent on on-line databases. The greater use of TEXTLINE by end-users can be explained by the fact that, unlike the Library staff, the researchers had no other external database hosts to choose from, neither did many of them have access to the press cuttings collection in the Members' library (two-thirds were barred from using the library). All their needs for current affairs information was being channelled into a single host, TEXTLINE. However, the fact that just one terminal, a long way from most people's offices, could generate such a volume of searching says an awful lot about the potential end-user appetite, especially for that premier commodity - news.

Monthly use was remarkably uneven ranging from a little more than 22 hours in October to a peak of 70 hours in February. (POLIS session data also peaked in February confirming that month's position as the busiest on-line month of the year). What was very noticeable was the huge swings in monthly usage - they averaged 33% (Table 4.28). In only one case - May - was the swing below double figures. TEXTLINE use was thus heavy, but also extremely variable - as POLIS' was.

TABLE 4.27
USE OF TEXTLINE: CONNECT TIME (Logs)[1]

Month	Library staff	Research assistants	Total
Oct	14.7	22.6	37.3
Nov	10.3	39.9	50.2
Dec	6.2	29.5	35.7
Jan	9.2	38.2	47.4
Feb	13.1	70.2	83.3
Mar	13.7	56.9	70.6
Apr	n.a.	69.2	-
May	n.a.	62.6	-
Jun	n.a.	53.6	-
Jul	n.a.	62.5	-
Aug	n.a.	32.5	-
Sep	n.a.	36.3	-
Total	132.6[2]	573.9	706.5

NB. n.a. = not available. 1. Connect time in hours. 2. Estimate, incorporating a figure for the period based on use of other on-line systems during that time.

Both groups of on-line searcher used TEXTLINE variably but end-user searching was more volatile, wildly see-sawing from month to month. Most noticeably, the peaks and troughs for Library staff were far less pronounced. Also the rises and falls in monthly use of the two groups did not always correspond: in November Library use dropped by 30%, whereas that for Research Assistants rose by 77%. In March the situation was reversed with Library use rising (by 5%) and that for Assistants dropping (by 19%). Comparing tables 4.25 and 4.27 shows that, while the time spent on-line by end-users see-sawed erratically from one month to another, the number of searches conducted remained relatively static. Thus from April to July search numbers varied by only 12, whereas connect time varied by 16 hours. Also during that period connect time dropped by 9%, but the number of searches made actually rose by 5%. It would seem that Research Assistants were doing shorter searches during those months. Perhaps, they were just busier during that period, still needing to search but not with so much time to spare, or there were bigger queues then, and the ten minute rule was being enforced.

TABLE 4.28
USE OF TEXTLINE: RISES AND FALLS
IN MONTHLY USE (Logs)[1]

Month	Library staff	Research assistants
Oct	-	-
Nov	-30%	+77%
Dec	-40%	-26%
Jan	+47%	+29%
Feb	+43%	+84%
Mar	+5%	-19%
Apr	n.a	+22%
May	n.a	-9%
Jun	n.a	-14%
Jul	n.a	+17%
Aug	n.a	+48%
Sep	n.a	+12%

1. Connect hours

Given the constraint of sharing one terminal with dozens of other Researchers - and a terminal that was a long distance from their offices, the volume of end-use of TEXTLINE can only be regarded as phenomenal. Driven, no doubt, by an all-powerful need for current affairs information. Whatever statistic you take: the peak number of searches conducted in a day (34); the 2340 search sessions conducted in a year; or the fact that end-use of TEXTLINE accounted for more than half of all commercial on-line searching done at The House of Commons; the performance was impressive. End-user searching was however volatile with the volume of searching varying enormously throughout the year. Library searching patterns, as with POLIS,

were different.

4.4.3.1.4 Length of Searches

The amount of time Research Assistants spent on a search would have come as something of a surprise if we had not been alerted to this by the POLIS dial-up data: more than half of their searches were accomplished in 10 minutes and an astonishing third were completed in 5 minutes.(Table 4.29). This was doubly surprising, as TEXTLINE was not generally noted for its speed at the time (Nicholas and Erbach, 1989; Chapter 1): response time could be poor, especially when searches were not qualified by a field label (and given the researchers' comments about codes - see Section 4.4.4 - they probably did not use them) and the menus, unless bypassed, were very cumbersome. Possible reasons for the relative brevity of the searches (negative ones first) were:

1. the searches were poorly formed and nothing was found - with nothing to print out searches were completed very quickly (a view Branch Library staff would subscribe to);

2. Research Assistants were not always aware of TEXTILE'S business bias, and they frequently failed to find the current affairs data they were typically looking for. They either obtained a poor return or nothing;

3. with the terminals almost in constant use and queues ever present, this put a constraint on searching time.

It was probably a combination of all these factors.

On the other hand, researchers may have been undertaking very straightforward searches (perhaps trying to trace a particular article) or were very happy with the results of quick and dirty searches - satisfaction with searching was generally high (see section 4.4.4).

Substantial searches were, nevertheless, conducted: 1 in 7 were longer than 30 minutes in duration. But, the pattern was very much one of lots of very quick fact-finding searches and a limited amount of in-depth research. A comparison with the length of searches conducted on POLIS (Table 4.4) discloses some interesting similarities. In the case of POLIS dial-up users, with whom they had most in common, TEXTLINE searchers took almost the same time to conduct their searches as POLIS users: in TEXTLINE's case 54% of searches were accomplished within 10 minutes - the equivalent figure for POLIS being 56%.

TABLE 4.29
USE OF TEXTLINE:
LENGTH OF SEARCHES (Logs)

minutes	Researchers		Librarians	
	n[1]	%	n[1]	%
0-5	818	35	163	41
6-10	441	19	89	23
11-20	501	21	87	22
21-30	245	10	30	8
31-59	264	11	23	6
60+	71	3	3	<1
TOTAL	2340	100	395	100

[1]Number of Searches

For Library staff TEXTLINE sessions were characteristically very brief with, over 41% conducted within five minutes and 64% within ten minutes. They were not, however, as brief as their searches of PROFILE' (61% within five minutes) and this might have something to do with both TEXTILE'S slow response time and the fact that staff were less concerned about cost considerations when searching TEXTLINE because of the lump-sum unlimited use arrangement arranged with the company. Library staff completed their TEXTLINE searches more quickly than Research Assistants with nearly two thirds of Library staff completing their sessions within 10 minutes, whereas just over a half of all Research Assistants did so.

Interestingly, the computer monitoring and the survey were at odds over the time researchers spent on-line to TEXTLINE. The survey showed that only 8% admitted to completing searches in 5 minutes (Table 4.30). Was it simply a case of users not realising how quick their searches really were or did users remove their failed searches from their calculations (and minds)? Perhaps, too, Research Assistants sought refuge in the *varies greatly* category - 17% of Assistants ticked this category. While researchers said they searched POLIS more frequently than TEXTLINE, when they did search TEXTLINE they spent more time on it. Thus while 8% managed to complete a search within 5 minutes on TEXTLINE, 18% of POLIS users said they completed searches in that time. This must be because it was a one stop service: TEXTLINE supplied answers to questions whereas POLIS simply provided bibliographic signposts to documents. Also, the greater time taken to complete a TEXTLINE search could possibly be explained by the fact that the system is (notoriously) slow to search and there is that much more data (i.e. text or abstracts) to print out.

TABLE 4.30
DURATION OF RESEARCH ASSISTANTS
TEXTLINE SEARCHES (Survey)

	Research assistants	
minutes	n	%
0 - 5	3	8
6 - 10	13	36
11 - 20	13	36
21 - 30	1	3
Varies greatly	6	17
Total	36	100

Quick searches - those completed in less than ten minutes - appear to be the hallmark of much remote on-line searching - and even when there are no hourly connect charges to worry about. Library staff accomplish their searches significantly more quickly than end-users. When questioned end-users thought their searches were longer than they turned out to be in practice.

4.4.3.2 File selection

At the time of the survey TEXTLINE offered four files for searching: TEXTLINE (abstracts or full text of periodicals and news stories); Newsline (a companion, updating file to TEXTLINE, offering the most recent weeks' expanded headlines); Dataline (a company accounts service); and Country Briefs (which provided reference type facts on countries around the world). All four files were used by Research Assistants, but TEXTLINE accounted for almost all the use - 97% of their time on-line was spent in this file (Table 4.31). Country briefs was accessed for some 6 hours - although a few long searches basically accounted for this total. Newsline and Dataline were used for around two hours, with searches being characteristically very short, usually of less than 5 minutes in duration. In the case of Library Staff, too, the TEXTLINE file itself accounted for the vast bulk of use. Newsline obtained light use but no use was made of Newsline or Country Briefs, the company information file. It would seem that Library staff had no need for detailed company accounts or country profiles - they would have obtained this data from their extensive reference collection.

TABLE 4.31
TEXTLINE FILE USE (Logs)

	Textline		Dataline		Newsline		Country	Briefs	Total	
	Lib	Res	Lib	Res	Lib	Res	Lib	Res	Lib	Res
n[1]	66.8	246.8	-	2.3	0.5	2.0	-	6.1	67.3	257.3
%	99.2	95.9	-	0.9	0.8	0.8	-	2.4	100	100

1. In hours.

Both groups of on-line searchers primarily concerned themselves with the main TEXTLINE file, though it was the end-users who in fact used the full range of files, and not the Library staff. This was probably because the Library staff could find the information contained in files elsewhere, whereas the end-users had very little choice.

4.4.3.3 Types of search conducted

Not surprisingly TEXTLINE was used rather differently to POLIS: for instance, Research Assistants would use it to look up a press item on food irradiation or check the number of times Mrs Thatcher had talked in the press about short-range nuclear weapons. According to the survey (Table 4.32) well over half of the TEXTLINE users (56%) looked up social/political information (e.g. fish farming, government defence contracts) frequently or very frequently. One-fifth (21%) of them also made frequent use of business/economic information (e.g. Monopolies Commission on Breweries). Little use was made of company (the low use of Dataline confirmed this) and EEC information: 85% having seldom or never searched for this type of information. As suspected, TEXTLINE appeared to being used for general current awareness - for broad sweeps of a subject, for which it was not really suited, rather than for its company and financial information, for which it was exceedingly well-equipped. A (slightly jaundiced) Library staff member explained: *most researchers think newspapers and for them, TEXTLINE makes it easier to look through newspapers; someone else is doing the work for them - they're very much into that.* American researchers particularly favoured the broad-sweep - allegedly Members deliberately gave them very broad subjects to research - to keep them occupied.

TABLE 4.32
TYPE OF SEARCH CONDUCTED ON TEXTLINE
BY RESEARCH ASSISTANTS (Survey)

	V. frequently		Frequently		Occasionally		Seldom		Never		Total	
	n	%	n	%	n	%	n	%	n		n	%
Company search	-	-	1	3	4	12	25	76	3	9	33	100
General economic search	1	3	6	18	11	33	14	42	1	3	33	100
Social/political search	6	18	13	38	13	38	2	6	-	-	34	100
EEC Search	2	6	3	9	6	18	21	62	2	6	34	100

Library staff mostly used TEXTLINE in a much more specific and targeted way - they used it occasionally for company searches, company personnel biographies and for locating articles in newspapers and journals not indexed by the Library, or covered by PROFILE.

End-users were not searching TEXTLINE on its coverage strengths: they were using it as a catch-all newspaper service, which it was not. The better provided Librarians used it as a niche service.

4.4.4 User satisfaction and problems experienced

Given what has just been said, it was not surprising to find that, amongst Research Assistants, levels of satisfaction with TEXTLINE were not as high as they were with POLIS, though they were still relatively high (Table 4.33). 72% of TEXTLINE users said that they were satisfied with the results of their searches, while another 25% said that they were satisfied some of the time. This finding was very similar to that of Jacobson and Ullman (1989). It does though conflict somewhat with the views of Branch Library staff, who observed the products of many TEXTLINE searches ending up in the bin.

TABLE 4.33
SATISFACTION WITH TEXTLINE SEARCHES:
RESEARCH ASSISTANTS

Usually	Sometimes	Not normally	No answer
26 (72%)	9 (25%)	0 (0%)	1 (3%)

Over half the TEXTLINE users questioned said that they occasionally experienced problems when searching the database, but only 11% frequently experienced problems (Table 4.34).

TABLE 4.34
EXTENT TO WHICH RESEARCH ASSISTANTS
EXPERIENCED PROBLEMS WITH TEXTLINE

Frequently	Occasionally	None	Cannot remember
4 (11%)	20 (56%)	10 (28%)	2 (6%)

Unexpectedly, the chief problem was not the shared access to the single terminal, but formulating a search enquiry - 36% of respondents singled this out as being a problem area (Table 4.35). In fact, access problems - 7 or 8 researchers jostling for the same terminal - was ranked third, after the menus, about which a quarter of respondents complained. Typical of the complaints voiced was this one : *the TEXTLINE menu is particularly vague and difficult to understand.* Users also experienced other problems like equipment failure - the TEXTLINE on-line station broke down regularly. According to one of the Commons' librarians when it did it threw the American researchers into panic: *they were horrified that their information lifeline has been*

taken away. There was also the slow and limited printing facilities - only a screenful of information could be printed at a time - though most people appeared to have a resigned acceptance about this. A number of users experienced difficulties with the TEXTLINE codes: *TEXTLINE was a nightmare because of the codes* (TEXTLINE has codes for companies, sources and much else).

TABLE 4.35
NATURE OF PROBLEM EXPERIENCED WITH TEXTLINE
BY RESEARCH ASSISTANTS (Survey)

Problem	no of times problem mentioned	% of respondents experiencing the problem (n=36)
Poor access	8	22
Equipment failure	4	11
The menus	9	25
Formulating a search	13	36
Printing	3	8
Response time	1	3
Total	38	100

With a range of other information systems to hand (e.g. PROFILE, DIALOG BLAISE etc.), Library staff were not so impressed with TEXTLINE. The Library Reference staff thought it to be too much of an end-user system - slow and cumbersome, but Library research staff, somewhat closer to being end-users themselves, had a higher regard for the service and were struck with the simplicity of the service (i.e. its menu-driven format), and wondered why the House of Commons could not do something like that for Parliamentary Questions. Library research staff, especially those from the Economics unit, appreciated its business orientation - it was really the only specialist business host that the Library subscribed to - and liked the results obtained from, what one Staff researcher, dubbed its *scatter gun* approach to information retrieval (this was probably an allusion to the high recall obtained from the system compared to the much more precise result normally obtained from POLIS).

TEXTILE'S coverage (or more accurately its lack of it) also came in for criticism. It was the way that TEXTLINE selected and rationalised stories that staff were not too happy with. This lead to *too many wasted sessions on TEXTLINE*. Staff here were referring to the fact that TEXTLINE did not take 'general' news items (essentially, only business stories were 'cut' from the newspapers it covered). Neither were news items duplicated, so a 'national' story (covered in the national press), say, from the *Glasgow Herald* would not be covered, although, of course, Glasgow Members might well be interested in the *Herald* story to see what viewpoint or angle was taken.

End-users were generally satisfied with their TEXTLINE searches, but they did encounter problems with the service - more so than with POLIS. With a smaller user base, access was not the principal problem, but formulating a search was. However, the problems they experienced did not stop them using TEXTLINE heavily. Library researchers were more pleased with TEXTLINE than their professional colleagues.

4.4.5 Training

As was the case of POLIS, Research Assistants were given only a 5 minute basic introduction to TEXTLINE during the Library tour. The true value of this can be seen by the fact that 83% of the Research Assistants surveyed felt that they had obtained no training at all in the use of TEXTLINE. Yet, almost half of the TEXTLINE users expressed an interest in being trained. This was in fact a lower proportion that wanted POLIS training - Research Assistants obviously thought that TEXTLINE was easier to search.

TEXTLINE was largely used by untrained users. The majority, however, wanted to be trained.

4.5 Delegation (Both systems)

Despite the fact that the House of Commons worked on a highly delegated form of information provision - doing actual research for its users, offering oral and written advice as part of the job - the vast majority (70%) of MPs' Research Assistants favoured doing their own searches. The reasons were very much those given by the journalists at *The Guardian* (see section 5.5), though the preference to do it for yourself was markedly more stronger at The Commons. For most Research Assistants searching on their own offered *more control and flexibility.* During the session, they could formulate their own search, broaden or narrow the search enquiry at will or, maybe, follow an interesting area of investigation. Then there was the assessment of relevance argument: a Library staff member would not be able to *appreciate the significance of information,* whereas the end-user could *assess the information and not miss anything.* (It was interesting to hear, what was, after all, another kind of intermediary say this). Moreover, the Assistant may have a number of searches to do or a number of lines of enquiry to carry out - and in these circumstances it was rather more difficult to delegate.

Of course, delegation only works when there is someone to delegate to. One Research Assistant mentioned that it was sometimes hard to get assistance: the Library staff

were often hard-pressed for time, and they would satisfy the demands of the Members first, though, of course, Research Assistants were searching on Members behalves. Only a tiny minority (7%) of researchers preferred to delegate the search to the Library staff. One of them said that Library staff were experienced and helpful and could do the search more quickly. In addition, some of those who did their own searches, would ask for assistance from the Library staff if they could not find what they wanted, or if there was a long queue at the terminal. The remaining 12% of Research Assistants expressed no preference at all. According to one respondent *it depends on what the search is for*: he would handle the easier searches and ask the Library staff to do the more difficult ones. Another said: *they* [Research Assistant and intermediary searching] *are not mutually exclusive and there must be a balance between the two.*

Why did Research Assistants then feel so strongly about doing their own searches? To answer the question it is necessary to view the broader information seeking behaviour of politicians. Politicians basically require two types of information: (1) neutral, factual type data; (2) politically slanted data, data, that reflects the viewpoint of the politicians' party. Generally, the Library supplies the former - indeed, to ensure its impartiality it studiously avoids getting involved with the latter, and the retrieval of politically biased is left to the researcher, who inevitably are of the same political persuasion as their Member, and consequently in the best position to make judgements about the political worth of the data.

Despite the rich resources that were made available to them - and there can be few Libraries in the UK as well-resourced or as pliant as that of The House of Commons, MP's researchers showed a strong preference to search databases themselves.

4.6 Comparison between POLIS and TEXTLINE

End-users of both on-line systems were asked to comment on their relative usefulness and user-friendliness. Nearly half the people who used both systems said that POLIS was more useful than TEXTLINE - and this was demonstrated in the connect hour figures, though not in the session data (possibly it would have been different if they had access to PROFILE instead). A quarter, however, thought that they could not be compared as they were used for different purposes: POLIS, for parliamentary information, and TEXTLINE for current affairs information. A larger proportion favoured POLIS because it contained Parliamentary information, which was directly relevant to their work whereas TEXTLINE, as one astute respondent said, was *a*

commercial database catering for commercial users. In fact, TEXTLINE did not really fit the bill. PROFILE would have been a more suitable choice as it contained a wider range of current affairs information and was more full-text in nature.

More users (46% as opposed to 29%) found that POLIS was easier to use than TEXTLINE. One said that *the menu [of POLIS] was easy to follow*, and another thought that POLIS was a *more interactive system.* (Did they mean it provided more on-line help?).

4.7 Conclusions

• The House of Commons study featured a group of users - Research Assistants, who, though political by qualification, experience and the nature of their work, could be considered intermediaries themselves. None were, however, professionally qualified in information work or even trained in the most basic techniques of on-line searching. Unlike the information professional they were mainly, but not always, searching on behalf of just one individual.

• Research Assistants constituted a large and very active group of on-line users at the House of Commons. From the comments books in the Branch Library which listed POLIS and TEXTLINE users, it could be gauged that there were about 100 in-house end-users at The Commons. Additionally, twenty-two had dial-up facilities to POLIS in their offices or constituencies. The membership of the two groups did overlap, but that probably still left around 115-120 Research Assistants who were actually engaged in on-line searching - about 20% of the potential end-user population. By any count then, they represented a major force of end-users. Indeed, at the time, there were as many end-users at the House of Commons as at Rupert Murdoch's News International newspapers, that centre of end-user searching in the UK. **Large numbers of end-users can exist even in an organisation that has Rolls Royce-like delegating facilities.**

• Research Assistants obviously took to on-line searching with considerable enthusiasm - and they were keen to do it themselves. This was reflected in the very high volumes of on-line searching undertaken. Taking POLIS and TEXTLINE use together, researchers spent 1752 hours on-line during the year surveyed. Perhaps that figure does not appear unduly high when given a potential user group some 600 strong, but as only 120 or so of these were active in an on-line sense this does represent 14 hours spent on-line per person per year. These figures are even more impressive when the date and relative infancy of Research Assistant searching at the House of Commons is taken into account. Dial-up facilities had only been available for about two years when the study began - and really not promoted at all. Even the Branch Library terminals were no more than four years old - and even these facilities were not heavily promoted. **End-users can clock up high volumes of use.**

• In the particular case of POLIS, whatever measure of use was taken (connect time, search sessions, records displayed), Research Assistants made extensive

use of the system and were highly appreciative of it. Their use of the system was put in the shade somewhat by the Library staff, but considering that: they were untutored users using a system that was poorly promoted or supported; and the constraints under which they searched (the vast majority had to share a single terminal, which for many was hundreds of metres away from their places of work), one must regard, for instance, 96,804 records displayed in a year to represent high levels of use. Assuming an on-line population of some 120 research assistants, then each researcher was 'pulling off' of the system around 800 records per year. **End-users are strongly motivated to search and can overcome considerable barriers placed in their way.**

- High volume of POLIS searching, yes; but what of its sophistication and quality? That is much more problematical. Compared to Library staff their searches did produce less records and were more likely to result in a zero result. They also employed fewer command and their searches featured fewer fields. Nevertheless the questionnaire survey showed that few were dissatisfied with the results they got from on-line searching. Perhaps, this indicates that on-line searching was peripheral rather than central to their needs - an extra which might possibly throw up something useful or a means of confirming information already held and received informally from elsewhere (if that was the case, they were very much like the Library Research staff then). Or, maybe, with the Library access problems they experienced, that they were simply very grateful for anything they could get quickly. Given the nature of menu-driven searching, searches were probably inevitably going to be more simple: even so there was evidence to suggest that only a narrow range of approaches, options and commands were employed, which might not be so much a function of end-user searching but an inevitable outcome of untrained searching - really no one obtained anything that a professional librarian would recognise as training. **End-users did search simply but this could have been as much a function of the system they searched - a menu system, and lack of even the most basic training.**

- For the researcher one of the most surprising things to have come out of the study was the great differences in searching behaviour between dial-up and in-house end-users. While some of these differences - especially those involving use - can be attributed to the greater accuracy of the dial-up logs, what cannot be written off quite so easily is: the different times the two groups searched, the fact that dial-up users tended to concentrate their searching on a single file, their heavier use of controlled language searching, the narrower range of fields

employed in their searches and the greater propensity for their searches to end in no displays of documents. **There were often bigger differences in searching behaviour between in-house and dial-up end-users than between end-users and Library staff.**

• What, then, of the specific differences between Research Assistants' and Library staffs' use of the on-line facility? Well, as previously mentioned there were differences at the general level, but one could always find a group of Library users with which they had something in common, and, indeed, it was the similarities in searching styles that were most noticeable. Take the use of the Research Assistants' Branch Library terminal. Research Assistants there conducted more on-line sessions than three Library Research units and nearly as many as the Reference Room staff, who had two terminals to the Assistants' one; in regard to connect time, while they spent just a sixth as much time on-line as the Library's busiest user (the Oriel Room), which had three terminals, they spent more time on-line than two Research units; in terms of records displayed, they had six Library user groups behind them and displayed more than twice as many records as the Reference Room staff; and their productivity rate (records displayed divided by connect time) of 125 records per minute was above par for Commons groups. Turning to file choice, we find that Research Assistants spent less time in the CURRENT database than Library staff. When it came to search approaches, Research Assistants in the Branch Library: did have a marked preference for default searching (though just two percentage points separated them from the Library staff of the International Affairs section); they showed a strong taste for the form search (no doubt heavily influenced by the menus); they were less inclined to employ thesaural (SIT) terms in a search (again this can be attributed to the menu: their dial-up colleagues, who were not so constrained, used SIT terms more frequently than Economics and International Affairs); and they used less of POLIS's search features (e.g. truncation and the thesaurus). **End-users did not always conform to their stereotype and in many areas of searching they shared similarities with groups of Library searchers, especially the Library's subject specialists with whom they shared similar subject and academic backgrounds.**

• One's impression of the on-line statistics is how different they are. There seemed to be not a lot in common between the various POLIS searching groups. Statistical support for these observations were sought. A chi-square test was run against possibly the monitoring statistics' most significant dataset - the fields

people searched. Indeed, the chi-square test lent substantial weight to our observations. When we compared pairs of on-line groups, all groups showed large differences between them; the smallest differences occurring - not unexpectedly - between the two Research Assistant groups (Branch Library and dial-up users).

- A stereotypical picture of a library professional searcher would portray them as controlled language and publication form searchers - yet the POLIS data collected here suggests that in fact they are not that dissimilar in this to end-users. While they were bigger users of the controlled language, this resulted largely from the menus which deprived users of the approach - dial-up users, who had a freer choice, came very close to the Library figure. And as regards form, end-users exceeded the Librarians in the use of the approach, but again this might have been influenced by the menus which favoured the form searcher. **Stereotypes do not always fit, especially when menu-driven systems are being considered.**

- The menu-driven approach to searching POLIS, designed to encourage end-user searching, generally supported the form of publication search. However, the original menus were misconceived and the assistance they provide misdirected, because the vast bulk of searches were conducted on subject : subject accounted for 72% of all the approaches taken. And it was the default subject search that was largely practised. Even amongst the Library staff the default subject search was preferred; the huge investment made by The House of Commons Library in its controlled language (SIT terms) if felt at all, was largely felt indirectly. Clearly, more needed to be done to provide assistance to the subject searching Assistant. The reason why end-users were held in such poor regard by Library staff could well be because they received little help in the way of subject searching, indeed they received very little training of any kind. Researchers identified the formulation of the search query as their principle on-line problem in the case of TEXTLINE, and second greatest difficulty in the case of POLIS. The Library did have a problem in this connection for they felt that they first had to identify Assistants who are worth training; that is, those who were likely to stay for a sufficient period to justify the training investment. **End-users are mainly subject searchers and as such they experience problems in choosing what terms to search with. Author searching is notable by its absence.**

- On-line searching by Research Assistants was a very variable and volatile activity (not unlike Politics in this respect), tied most firmly to the work rhythms of the House, peaking when The House was at its busiest. Most Research Assistants searched irregularly (though not infrequently); use varied considerably throughout the year, throughout the week, but less so during the day. **End-user searching varied significantly in volume from day to day and month to month and its peaks and troughs were not the same as those of the Library staff.**

- What of the quite astonishing amount of use made of TEXTLINE by Research Assistants. Even given an on-line service that far from met all their information needs - few used TEXTLINE for its business coverage, a well-stocked library and the opportunity to delegate the search to Library staff: Research Assistants used TEXTLINE in large numbers and large volumes. Their use far exceeded that of the Library staff; in fact, it accounted for an astonishing 55% of **all** external on-line searching conducted at the House of Commons. Indeed, what all this has shown is that politicians, like journalists, were capable of registering higher volumes of on-line use than the intermediaries who served them. Undoubtedly, the pressing need for current affairs data contained in newspapers combined with the restricted access to the Library's cuttings collection, provided much of the incentive. It was a shame that much of it was misdirected for TEXTLINE was not the most appropriate host - PROFILE was), though, the fault was not theirs. **Researchers had a cavernous appetite for current affairs data and they showed this by their heavy use of the only current affairs system they had - even though they recognised it was not the most appropriate. End-users are not necessarily single Host-bound: two very different services were embraced by Research Assistants - the large majority searched both.**

- The differences in searching behaviour between the two groups of TEXTLINE searcher was very much the same as it was with POLIS: (1) though typically quick searchers, end-users did take longer over their searches than Library staff; (2) end-user searching was more volatile - they searched heavily at different times to information professionals; (3) end-users made greater use of all TEXTLINE's files; (4) Research Assistants were generally more satisfied with the service, than Librarians but then with just one current affairs service to the Library's three - PROFILE, DIALOG and TEXTLINE, they had to be. **There were many similarities between TEXTLINE and POLIS searching.**

- As to the quality of TEXTLINE searching? Well, this takes us back again to the question of perception and who do you believe, for Library staff were of the opinion that much end-user searching was of a highly questionable quality and much of it being recognised as such by Research Assistants - and, hence, binned. However, TEXTLINE was obviously an information lifeline for many of the researchers and as long as it provided some flotation nobody minded - or indeed noticed - how well they floated. If six feet of printout had to be dispensed with in order to get the information that was required - something it has to be said was not proved, then, maybe, that was a small price to pay for what Research Assistants would regard as information on-tap. Also, 'free' unlimited use on-line facilities almost encourages a wasteful, throw-away culture. Furthermore, Library staff would undoubtedly be more critical of TEXTLINE because they had alternative services with which to compare it with. The inappropriateness of much of TEXTLINE's coverage led to lower levels of satisfaction with the system - as compared to POLIS. **There is a strong difference in opinion as to the success of end-user searching, with end-users believing that they are largely successful, and librarians begging to differ.**

- Research Assistants made a number of suggestions for improving their on-line lot. Access was their principle concern: to more databases and on-line facilities in their offices. One Research Assistant, showing commendable knowledge of on-line sources, wanted access to the JANET network. The real problem with this suggestion for the Library was how to monitor end use. Unlimited use and relatively cheap on-line arrangements were required if unregulated on-line searching was allowed. CD-ROMs may provide a solution in the future - see the Postscript section (4.8). Two respondents wanted networking, with direct link-up to Member's offices, something that has at last come about (House of Commons Information Committee, 1994). **The end-user appetite remained undiminished by the experience of searching POLIS and TEXTLINE.**

- Most of the complaints made by Research Assistants were quite positive: what they wanted was an improved on-line service. They clearly liked what they got from on-line databases and wanted an improved and more comprehensive service. Put simply, they wanted lots more on-line facilities. What was clear from the range and quality of the complaints was that many end-users - contrary to expectation - had a high appreciation of the problems encountered when on-line searching: not all end-users are necessarily naive users. Patently though, their needs and aspirations (and those of their Members) were not being fully

met by the existing facilities. The problem really lay with the fact that their requirements were not highly prioritised by the Library (or indeed, by The Commons): Members of Parliament were the Library's primary priority - and they were happy to delegate to the Library. No one had the corporate responsibility to ensure that Research Assistants were given the necessary tools and the training in the use of these tools. However, the question one must ask is whether the needs of Members and of their information assistants can really be distinguished? The answer must surely be no for they, together with the Library staff, were part of a very complex and close-knit, overlapping information community. **If end-users showed lower levels of search skill, they were alert to the general problems associated with on-line searching.**

- And to the culture of on-line searching at The Commons. Despite spending vast sums of money on maintaining and searching on-line databases - perhaps as much as a third of the Library's budget was so spent - the Library did little to promote on-line databases to Members and their staff or support what limited facilities that were provided. In fact, there was not a little 'secrecy' connected to their use of on-line databases: terminals in the Oriel Room were deliberately tucked out of sight, below the counter line, so Members were not rudely confronted by them, and Library staff did have a strong preference for conducting the on-line search when the Member was not around, or not looking, which is hardly 'textbook' searching. How, then, do you convey to members and their research staff that a computer terminal, taking up less space than half a dozen books and functioning as the Library's catalogue most of the time, can provide a window to a collection of full-text journals that far exceeds the Library in size? Well, a lot of Research Assistants find this out by accident - and a lot do not of course. There was also a feeling among Members (and Library staff, too) that because the Library was so all-powerful, possessing *those black boxes* that they could answer anything. But such a view neglects the fact that a lot of information seeking is triggered off by browsing, something which is not easily delegated. **For the Library end-use was more a safety or overflow valve, it was not part of an overall integrated information policy.**

- Attitudes towards on-line differed amongst Library staff. The research activities of the Library were highly prized by Members; yet ironically, the Research Division, who were arguably the Library's most versatile handlers of information (they were active in the retrieval, processing and packaging of information), were least touched by modern developments in information handling, and in

some cases, antagonistic towards it (or some of its manifestations, like POLIS). To the question *Have on-line databases resulted in any improvement to the service you provide for Members?*, many, but not all, Library research staff answered *no* - despite having access to on-line services that would seem to meet their every need (POLIS for parliamentary information and PROFILE for current affairs data). We hardly ever came across on-line databases being used in any research capacity; for instance, in the way that some journalists used PROFILE to piece together a story (see section 5.4.5). Computer terminals were, of course, readily available to all, but when it came to using them research staff seemed quite ready (as Members were) to distinguish between the twin (research) activities of searching for and processing information: they saw on-line databases as a vehicle for the former activity and delegated the retrieval task to the Library professional attached to the section. In such circumstances, the delegation model becomes a complex one and one surely fraught with danger: the Member delegates to the Library research staff member and s/he then delegates to the Library professional, who eventually does the on-line search. **Not all Library staff were enamoured with on-line systems, POLIS generated the greatest ire amongst members of the Library research units.**

4.8 Postscript

Towards the end of 1990, just as the main study period was coming to an end, it looked like the end-user 'bubble' was about to burst. Reuters decided to end all unlimited use contracts and, as a result the Library, albeit reluctantly, withdrew the TEXTLINE facility from the Branch Library. With that level of demand in the pipeline, the Library felt they could not afford to take on an open-ended commitment to pay £50/£60/£70 an hour (depending on volume me of use), especially when they had been paying the equivalent of about £15 an hour for TEXTLINE.

What would happen in the future was then not quite clear. In the new (Derby Gate) Branch Library, which was built for them 1991, Research Assistants would be able to ask Library staff to do PROFILE searches for them. However, it would have been impossible for Library staff to take on that huge level of TEXTLINE end-user searching. Members were undoubtedly going to suffer: you cannot lose 570 odd hours of on-line time and not notice it - and there were Questions in The House about its removal. The Library solution was to replace TEXTLINE with Newspaper CD-ROMs: no thought was given to the fact that MPs and their assistants prized currency and speed of delivery above everything else - and Newspaper CDs typically serve up

news four to six months old.

In 1993 events moved apace. That year saw the introduction and piloting of a Parliamentary Data and Video Network (PDVN). The pilot involved 105 Members and their Assistants. The provision of such a PDVN has been mooted for many years (House of Commons Services Committee 1984, 1991) and was at last becoming a reality. Providing MPs and their Assistants with direct access to POLIS, CD-ROM networks and commercial on-line services was very much at the heart of this development. A report on the pilot study, provided some interesting data on end-use (House of Commons Information Committee, 1994) and demonstrated that not many things had changed. Change has been slow: the loss of TEXTLINE had undoubtedly taken some of the momentum out of the end-user revolution.

Although 90% of Members' offices had now IT facilities of some kind or another, barely half the Members used the facilities themselves or were doing anything serious with them. IT was simply a case of secretaries word-processing correspondence. No strategic use of IT then.

Just 13% of members and their Research Assistants surveyed had obtained any IT training. It was - as it was in this study - the most common area of concern and worry (and nobody had done anything about it in the time since).

In regard to current on-line searching activities of politicians: 20% claimed to use POLIS. Now that is almost exactly the same proportion that searched POLIS in 1990 - no change there either. However, showing the value to MPs of rapid information delivery, the complement of MPs with dial-up facilities had almost doubled to 50. Also - something new, 6% of respondents now accessed Hansard on CD-ROM, and 5% FT PROFILE. Frequency of use data was obtained and is summarised in the Table below:

TABLE 4.36
USE OF ON-LINE SYSTEMS AT THE HOUSE OF COMMONS BY MPS
AND THEIR RESEARCH STAFF : AN UPDATE (1993-1994)

	Daily	Weekly	Occasionally	Never	No answer
POLIS	7	11	10	3	3
CD-ROM	5	9	4	9	7
FT PROFILE	0	3	1	0	30

Source: (Ibid., 111)

There seems to have been an increase in POLIS use: 23% now said that they searched the service weekly, whereas the figure for 1990 was 11%. CD-ROMs have obviously taken the place of TEXTLINE: 19% of respondents searched CD-ROM daily, a figure that drops to 15% if the people that did not answer the question are counted as non-users. Even at TEXTLINE's height only 3% of Assistants used it daily. So despite concerns over currency their has been an increase - the availability of the full-text Hansard (something that politicians had long argued for) probably boosted the figures. The use of PROFILE did not come from Members taking out personal subscriptions: no Member has done this yet. Access to PROFILE was, in fact, provided on a short term (12 months) and experimental basis. 13 members were signed up for the experiment and given training and a search allocation. The experiment proved disappointing with few Members taking up their allocation, leading the Library to the opinion that *most members are content for Library staff to carry out searches on their behalf* (Ibid., 99). Yet this was not the case with TEXTLINE. This conclusion is highly suspect given the small numbers involved, the fact that the questions were asked of Members rather than their Assistants and the difficulties some Members experienced in dialling out through the PDVN.

Of the IT facilities on the network, on-line services were not thought to be the most useful - in fact FAX was thought to be most useful. CD-ROM - Hansard and daily newspapers - were a close second. POLIS was thought next most useful by Members and FT PROFILE followed at a distance. There were complaints about the on-line systems, too:

- CD-ROM - *user hostile, so out of date it is useless;* the problems in searching CD-ROMs with different search languages came to the fore : *I cannot see people using the CD-ROM network if they are going to have to work with a manual* (Ibid.; 39)
- POLIS - *no full text facility, user hostile, very slow;*
- FT PROFILE - *dialling out is difficult.* (Ibid.; 112)

The whole issue of who should really be doing the on-line searching of external hosts (Research Assistants or Library staff) has not gone away, for there was some disagreement amongst MPs in the Committee's deliberations, with one believing that end-use would result in money being wasted - better to leave this to the Library, he thought, possibly sending their questions electronically to the Library via the PDVN. However, another respondent, understanding rather better the nature of MPs' information needs, felt the Members need to see and shape information in a political context, and this meant that facilities for end-use had to be provided (information

needs have not really changed over the time).

Finally, in regard to future use, courtesy of the proposed PDVN, half were looking forward to using POLIS and Hansard on-line, 35% the Press Association newswire (currency concerns, are prominent again) and 31% FT PROFILE (this throws further doubt over the validity of the PROFILE experiment). But really what they all wanted was a super all embracing system that had an enhanced POLIS (complete with the full-text of the Library briefs), CD-ROMs and PROFILE on it - all searchable by the same command language (Ibid. 330). Could this be Walton's (1983) *black box* ? (see page 30).

Summing up, end-use has been stimulated by the network and the presence on the system of full-text services. Currency and speed of delivery concerns are ever present, but still nothing has been done about training, and as a result searching styles probably remain the same. Certainly the people doing the on-line searching - the Research Assistants are very much the same, although it is possible to detect a greater understanding of the process on behalf of Members.

CHAPTER 5
JOURNALISTS AS END-USERS
CASE STUDY: *THE GUARDIAN*

On-line databases were first introduced to *The Guardian* in 1984, when the paper took out a subscription to FT PROFILE, then World Reporter. FT PROFILE was the first choice simply because of its full text news coverage. *The Guardian* soon appeared on it - and in so doing, becoming the first newspaper to appear on PROFILE (in fact the first UK newspaper to appear on any on-line host). Initially, the single terminal was placed in the neutral precincts of the managing editor's office (it was always intended to be used by journalists), but soon moved to the library where it was largely searched by librarians on behalf of journalists. Gradually a few computer-oriented journalists began searching the system themselves - increasingly from home, where they dialled in from their office-issued Tandy lap-top computers or on their own personal computers. However, it was not really until 1987 that journalists could search the system at their desks - via their ATEX editorial terminals, and it was only then that end-use really took off. This was about the same time as end-user facilities were introduced at The House of Commons.

The Guardian offered a rich test-bed for examining the on-line searching habits of end-users. There were three reasons for saying that. Firstly, searching can be studied over a very long period of time, 1984-1995. Secondly, journalists have always had the choice as to whether they searched themselves, delegated that task to the library staff, or, indeed, did both. Thirdly, cost has never really been a major barrier to end-use, because the royalties obtained from PROFILE from the use of *The Guardian* file has always offset the costs of searching PROFILE. Additionally, journalists were sheltered for some of the time by generous unlimited use lump sum agreements. Journalists had thus both freedom of choice in on-line matters and long and unrestricted access to, arguably, one of the world's most user-friendly on-line systems - FT PROFILE. There is a caveat to all this though, for while cost has not always been a major issue, *The Guardian's* management have always attempted to exercise control over use, and they have done this crudely by restricting journalists initially to just one password between them all - since 1994 the number has increased to two. *Guardian* journalists have had then, plenty of time to come to terms with on-line databases and, if they chose to search themselves, plenty of time to establish a searching style.

5.1 Research Methods

A variety of methods have been used to assess end-use at *The Guardian*. This was partly through necessity, but also in-keeping with the belief that it is only really from a mixed methodology that on-line searching behaviour can be fully described and understood. Although the study has been conducted over an extremely long time by the standards of information science research, the study cannot claim to be strictly longitudinal in nature. For, while the general line of questioning has always remained consistent and the researcher a constant, the population of journalists has changed over time - although not markedly so, except in the Features department. Also, a number of methods have been used at different times over the years. Nevertheless it is hoped that this study has inherited some of the virtues of longitudinal studies, like their ability to provide deep insights into change and it causes.

There were primarily four stages to the study, three of these stages being marked by the introduction of a new method of enquiry. To begin with (1985-1987) interviews and observation were the main research instruments. On-line searching was then barely established - an unknown to many journalists, and attitudes only slowly forming, therefore there really were no alternatives to these methods. Fourteen depth interviews were held with a selection of journalists representing the various departments of the newspaper. All these journalists had either expressed an interest in on-line or were already using the facility. Interviews lasted a period of one-hour and questioning was open-ended (see Appendix 6 for the schedule). The main problem encountered was not in getting journalists to talk, for they proved free and informative in these matters (see for instance the interview with Harold Jackson reproduced in *On-line searching* (Nicholas et al., 1987)), but in pinning them down to a date for the interview. Three interviewees were selected for a further half-day period of observation: both users and potential users were represented. Casual observation was conducted at the journalists desks: all information seeking and gathering was observed and noted, not simply interactions with on-line systems - with on-line in its infancy there was not many interactions. In addition six City journalists were observed at work for a period of a day (Appendix 7 for schedule). As they all sat in a reasonably small room they were all observed at the same time. The same topics were monitored. No problems were encountered in observing journalists at work: they were busy and self-confident enough not to be noticeably influenced by the observer. The eight library staff involved with on-line searching were also interviewed, as much for their perceptions of journalist searching as to determine their own searching techniques. Attitudes towards end-use and observations of journalists' on-line searching were obtained (see Appendix 8 for the interview schedule).

The second stage of the study - 1988-1991 - encompassed a period of slow, but steady, growth in end-use. This period was monitored routinely and less intensively. Regular quarterly visits to the newspaper were supplemented by repeat interviews with library staff: the pattern of questioning was much the same - as given in Appendix 9. In addition six more one hour information needs interviews were conducted with journalists off the premises of *The Guardian,* and again both users and non-users were represented (Appendix 7 for the schedule). The purpose of these interviews were largely to ascertain the extent to which on-line was changing journalists' information seeking behaviour. Information from the host involved - FT PROFILE, was also acquired: in the main this was 'confidential' data as to the volume of use and the experiences of PROFILE staff who had trained journalists. Stages 1 and 2 largely contribute historical and contextual data to the study.

By late 1991 end-use had expanded so much - overtaking that of the librarians - that Library staff though that the time was right to undertake a major survey of journalists use of on-line services - essentially FT PROFILE. This marked the third stage of the project. As a consequence, in Spring 1992 a questionnaire (Appendix 10) - a collaborative effort on the part of Library and researcher - was distributed to approximately half *The Guardian*'s journalists: 120 journalists - predominantly, but not exclusively, reporters, whose needs were felt to be more diverse and less obvious than those of the sub-editors - although this group was represented too. Questions concerned: volume of use, access problems, time of searching, retrospective searching, file selection, type of search conducted, use of commands, satisfaction levels, training and use of complementary information resources. After much badgering, 87 questionnaires were in fact returned, giving a healthy response rate of some 73%. In an attempt to get honest, critical and less-defensive responses, journalists were given the opportunity to remain anonymous, but in fact only two availed themselves of this opportunity. Library staff were also given a mini questionnaire survey - they were asked to answer one of the questions given to journalists - Question 8, which concerned itself with search success. Twelve questionnaires were distributed and eight returned.

In answering the questionnaire problems were experienced with the numeric scaling attached to some questions - it was not entirely clear as to which end of the scale (1 or 10) represented the most and the least; and also with the words used in the question framing - the term truncation was not widely known to journalists for instance. In some cases too, non-users answered questions that were really directed at users: the question (Question 6) which asked how far back in time journalists searched was a

good case in point. The data were analysed using SPSS: the Statistical Package for the Social Sciences. Since the survey journalists have continued to be interviewed (using the needs interview schedule) at a rate of two a year and close contact with library staff maintained, especially those involved in the on-line training of journalists.

It had been hoped that computer logs could be obtained from FT PROFILE to complement and underpin the questionnaire survey - as was the case with the POLIS study. Initially, the Host proved uncooperative, citing the market-sensitivity of the data, and it was not in fact until 1995 that the data was finally released - heralding the fourth and final stage of the study. Unfortunately, because of the delay, the data contained in the logs were not directly comparable, but nevertheless still valuable in their own right - they touch on areas not covered by the survey (number of documents displayed, for instance); offer the possibility of cross-checking the survey data, and provide an opportunity to see whether things had changed over the intervening three years. Computer logs for all *Guardian* password holders (journalists and librarians have their own passwords) were obtained for two day period 22-23 February 1995. Despite the fact that only two days data were made available, the logs provided details of 86 end-user search sessions and 99 individual searches, as well as a full record of the librarians' on-line searches for those two days (53 sessions and 69 searches). As journalists had to log in personally before they went on-line to PROFILE, it proved possible to pin down searches to individuals and departments. With just 39 identifiable journalists searching the system during the monitoring period the data was aggregated, rather than being sub-divided by department, job-title or gender, as was the case with the questionnaire data.

The data provided by the logs for each password holder were as follows: date, time and duration of session and search; files selected; search terms used, search and display commands used; fields searched; and number of records and screens displayed. In addition, by scanning the logs it was possible to identify input and spelling errors, and plot the development of search strategies. (A copy of a page from the logs, together with the data collection form, can be found in Appendix 11). It had been thought initially that the data would be analysed by SPSS, in the same way as the questionnaire data was, but the fact that: (1) there was a lot less of it; (2) the way it was presented by the host meant it was easier to analyse by hand; (3) it would have to be scanned manually anyway to detect strategies and errors; (4) disembodying the search through a statistical package would mean the interrelationship between on-line interactions would be lost; decided the matter, and

the analysis was done visually with the aid of word-processing and spread sheet packages.

Because PROFILE is an expensive dial-up host, its use was closely monitored by management. This meant that the logs were not problematical in terms of identifying search sessions and attributing data to individual searchers. So on the whole, while there might have been less data, they can generally be held in higher confidence than the POLIS computer log data. However, the logs did not come completely problem-free:

1. It was not always possible to pin down a session to a named journalist, although this only happened on seven occasions.

2. At busy times library staff sometimes used the journalists' passwords and had to be excluded from the analysis.

3. While it was always possible to mark the boundaries of a search session - dial-up and connect charges ensured this, it was not so easy to determine whether sessions involved a number of separate searches: fortunately most sessions were so simple and quick that this did not become a real problem.

4. No postings counts were supplied with the data, so search productivity data had to be obtained through other means - numbers of records displayed or screens viewed, and even then there were problems in establishing accurate figures.

5. While it was possible to detect input errors and inappropriate use of commands, it did not prove possible to (conclusively) detect poor search strategies. This was not a fault of the logs so much, more an inability to determine this given journalists' (sometimes) indirect method of searching.

5.2 The on-line system: FT PROFILE

FT PROFILE is a full-text current affairs host, best known for its coverage of the UK national and regional press. It now boasts 28 titles in this category, though this extensive portfolio has been built-up slowly over the years. Its coverage of UK press titles is unrivalled, and this together with the coverage of their own newspaper is its principal attraction for *Guardian* journalists. For convenience of searching British newspapers are grouped together in a file group called UKNEWS - this is a cross-file searching group which has much appeal. Foreign newspapers, wire services and weekly magazines, are other PROFILE sources that journalists are likely to be attracted to.

In its original command-driven form - the form *The Guardian* use, FT PROFILE has by common consent probably the fewest and simplest commands of all on-line hosts. As Harry Jackson, the *Guardian's* systems editor once said: *you can get through World Reporter* (FT PROFILE) *with the use of four words - and that is now reaching the upper limit of what people* [journalists] *can actually remember.* (cited in Nicholas *et al* 1987*).* Other factors which have contributed to making the system attractive to journalists are: (1) highly memorable command labels - **get, pick** and **select** for instance; (2) the ability to refine searches without resort to Boolean operators - allegedly the Achilles heel of end-users: the **pick** command serves instead, and is in reality a camouflaged Boolean **and**, working on the previous search set; (3) the intimate familiarity journalists have with the data covered (news stories); (4) the absence of potentially confusing bibliographic record; (5) natural language subject searching:- there are only a few files on PROFILE that have controlled language descriptors (another potential area of end-user confusion) - and they have only introduced the facility recently, and most of these files (with the notable exception of the *Financial Times*) were unlikely to have been searched by journalists; (6) the system works (almost) within the same time frame as the journalists - newspapers are at worst 24 hours late in appearing on the host, in fact the *Financial Times* is on the system before it is on the news stands.

5.3 The FT PROFILE users

There are approximately 240 *Guardian* journalists - mostly located at the Farringdon Road site, though some are situated in Manchester and others (the Foreign correspondents) are dotted around the world. Whether this figure represents the true potential size of the end-user group is not easily determined, for in common with many newspapers, *The Guardian* employs a lot of part-time staff, some on a short term basis and some on long-term basis - not all of whom work in the building. Currently the on-line and (library) users are predominately the hard-pressed Home affairs reporters; Foreign reporters who telephone in from abroad; the Foreign desk; the Features department and sub-editors attached to all these areas. City journalists (i.e. the financial ones), are relatively self-contained: they have their own databases (Datastream) and departmental library, and then there is the Sports department, which is composed largely of sports enthusiasts who seem to manage with very little help at all from databases or the library. Additional groups which are superimposed on these basic ones include specialist reporters, who have responsibility for specific subject areas (e.g. Medicine, Social Services) and are based predominantly in Home affairs,

and the smaller group of assorted managers: assistant editors, senior editorial management and departmental heads.

Home and foreign news journalists are the most populous and most demanding users of all: they require facts and background data very quickly, and usually at the last minute. Of the two sets of journalists, Home news journalist are under the greatest pressure - covering news on the ground as it breaks. The closer to home, the greater the pressure for rapid reporting: after all if we in Britain hear about an earthquake in Japan eight rather than six hours after it happened, it makes little real difference: but the announcement of an interest rate cut is quite another matter, and in these circumstances minutes matter. It is in the nature of general news reporting that journalists usually do not know exactly what they are looking for and frequently stumble upon information when they are looking for something else. It's the unexpected that the general journalist feeds on, and it could be argued that computerised information systems reduce the opportunities here.

Features journalists on the other hand generally have a narrower subject brief (e.g. *Arts page*, *Women's page*), know in advance the story that they are going to write and have (in theory) more time to research and write the story. As a consequence they have a prodigious appetite for information and are typically browsers - looking for ideas in cuttings files is a popular information activity. On the whole they are thought to be self-sufficient information seekers. There are though, general features reporters who write expanded feature length articles on current news stories: their needs are obviously somewhat broader. Foreign correspondents - *The Guardian* has around 20, some freelance - are of course geographically removed and remote - and frequently cut off from the international information flows (the Moscow and South African correspondents were recent cases in point). The correspondents need all the information that they can acquire, and in this respect, on-line can be an information lifeline for them.

The specialists journalists have, by newspaper standards anyway, one the narrowest of subject briefs (although perhaps not quite as narrow as the Sports journalists, who might specialise in cricket or rugby). They are by definition subject specialists and are in some ways similar to academics, in that they read the books and journals of their particular specialism in an attempt to keep up to date. As part of this process many have built up large, personal information collections, although like other journalists, they rely heavily on personal contacts for information.

Sub-editors - after Home news reporters the biggest group of journalists on the newspaper, probably experience the greatest difficulty obtaining information. They have a need to check, cross-check and verify information - like spellings and facts - literally within minutes and frequently when the library is unstaffed (late in the evening). Traditionally they have been poor users of the library. Because of the speed and urgency with which they work, they commonly fail to find what they want and, logically, it would be expected that on-line had much to offer to them. Sub-editors do have an editorial role in the paper - and indeed many departmental heads have a subs background. However, very few have been trained to search on-line. Indeed, hardly any *Guardian* journalist has obtained on-line training: training was offered at the very outset by the host, but few took up the offer, and of those that did some have now left the paper. We have here another group of untrained end users.

In addition to the journalists there were 13 library staff that searched FT PROFILE, most of whom were either professionally trained librarians or trainee librarians (of which there were three). They also searched PROFILE on behalf of *Observer* journalists and NEXIS on behalf of both groups.

5.4 Search characteristics

As was the case with the House of Commons study, which featured a survey and logs as well, the following analysis is potentially confusing, what with tables being based on different sample sizes. In respect of *The Guardian* study it is even more easy to become confused and care needs to be exercised when reading the tables generated by the survey and transactional log analysis. This is because the tables are built on different numerical bases. In the case of the questionnaire study, not only do you have the normal variation in base numbers that arises from different numbers of people answering a question - either as a result of a question not being relevant to them or, maybe, they simply chose to skip it through lack of time; but you also have people answering a question which they were not required or expected to. A good example of this was the question that asked whether journalists checked hard-copy resources before they went on-line - some non-users interpreted this question as meaning before they asked the library to do an on-line search. Then there were the questions that allowed for a multiple choice: the question that asked what types of search were conducted is a good example. In these cases the base number is typically larger than the number of respondents. Problems also arose with the logs - though they were different ones. In respect to logs the main source of potential confusion comes from the fact that not all the logs of individual searches provided the data that was needed.

A very good example of this was the number of documents or screens viewed in a search. Thus the number of searches for which this data was available was well down on the actual number of searches conducted. This was because, if the user had either used the **nobreak** command (in the case of screens) or the **all** command (in the case of document displays), there was no way of knowing what the outcome was. In recognition of the difficulties caused a reminder of variations in base numbers are posted beneath the tables.

5.4.1 Use data

What is of prime concern here is how much use was made of the on-line system. In the case of *The Guardian* this could be measured in a variety of ways: (a) by the number of search sessions conducted; (b) by the frequency of searching; (c) by the number of individual searches conducted; (d) by the length of time spent on-line (connect time); (e) by the number of records displayed; (f) by the number of screens viewed. The last two are really output measures, which can also shed light on other aspects of end-user searching (viz. search productivity and searching style), and as such are also dealt with elsewhere. Because *The Guardian* study covered the entire period of on-line at that establishment it was also possible to (roughly) plot the growth in end-use.

5.4.1.1 Growth in end-use.

In 1984, when FT PROFILE was first introduced into *The Guardian*, it was largely the library staff that used the system, though their number was boosted by a few computer oriented reporters, some City journalists, a medical correspondent and a few journalists who had to work from home because they lived a long way from the office. It remained that way for several years until 1987, the year when the ATEX computerised editorial system was introduced. FT PROFILE was quickly made available through the editorial system, and that was when end-use really took off. Proving - if proof was needed - that access and convenience were key aspects in the take-up of on-line searching. Speed of delivery is almost everything in journalism. What journalists particularly appreciated was the ability to download the results of searches into workspaces, edit them there and then incorporate the data into their own stories. Since 1987 end-use grew at a steady rate until a point was reached in 1992 when journalists and librarians searched FT PROFILE in almost equal volumes. Senior management appeared surprised (and worried) by the alacrity with which journalists took to on-line searching, and attempted to manage (contain) growth by

refusing journalists demands to have the number of passwords increased. Management also implemented a logging on system, by which individuals could be identified when they went on-line - hoping that this would act as a deterrent to going on-line too frequently. 1995 saw journalists' share of PROFILE searching increase to two-thirds, reaching a high water mark of 70% in March 1995. This does not convey the full magnitude of the end-user success story, because during this period library use also shot up.

There can be little doubt that, when journalists are presented with the opportunity to search themselves, they will do so in large and ever-increasing numbers. Interestingly, back in 1986 Harry Jackson, *The Guardian's* system editor, bemoaned the fact that so few journalists had taken up the opportunity to go on-line (Nicholas *et al.*, 1987). He referred to a phenomena that a number of commentators had noted at that time: an initial interest on being shown on-line for the first time and then a rapid waning in interest as the novelty wore off. So what changed things? Firstly, the improved access to the system already mentioned. Secondly, according to Harry Jackson, they got more used to *manipulating the ironmongery*. Thirdly, they were just that more aware of the technology: society is that more conscious of IT now, and virtually every newspaper now reports daily on some event connected to The Internet, E-mail, modems, CD-ROM etc. - and all this rubs off on journalists too, making machine-readable data much more acceptable and correct. Fourthly, the currency and coverage of FT PROFILE has improved enormously over the years, and as a result has become more and more useful to growing numbers of journalists. Fifthly, journalists increasingly saw the library staff answering their questions with a print-out or by consulting a terminal - and this undoubtedly has a promotional effect in regard to on-line use.

5.4.1.2 Frequency of searching and number of sessions conducted
The number of search sessions conducted by users and the frequency with which they search are not quite the same thing, although questionnaire studies in particular can be guilty of confusing the two. After all heavy use does not always imply very frequent use: it is quite possible that an on-line system might be used heavily, but not regularly, throughout the year - use being sporadic in nature. However, in the particular case of journalists it could be expected that frequency and volume were indeed closely related. This is because much of the work done in newspapers is very cyclic in nature - and very short cycles at that, typically 24 hours.

Survey

The questionnaire asked journalists how many times they searched PROFILE. It was left deliberately open in an attempt to avoid shoe-horning their responses into ill fitting frequency categories. For the purposes of the analysis the responses were grouped into four use categories - see Table 5.1A. According to the survey in 1992 just less than two-thirds (65%) of responding journalists searched FT PROFILE themselves. Probably, this figure is an over estimation, as it is quite likely that, amongst the non-respondents, non-users were over-represented. The incentive to return the questionnaire would have not been so great with this group, and journalists do have their pride and, although unlikely, might not want to have admitted to something which might have looked like an admission of failure on their part. Of those journalists that did search FT PROFILE, for 39% of them on-line searching was a daily activity - a level of activity not all professional searchers achieve. Certainly at a level sufficient to maintain skill-levels it would be thought. 18% of the journalists questioned were what might be classified as heavy users of the system, using PROFILE more than twice a day - six times a day in the case of one journalist. These levels are much higher than those found by Harman (1986) in her study of NEXIS-using Reuters journalists, but about the same levels as Jacobson and Ullman (1989) found. But even these relatively high use levels probably underestimate the potential to which PROFILE could have been used, for journalists were seriously constrained in their use of the system by: (a) a lack of available passwords - just two were available for the entire staff of reporters; (b) a lack of training and an absence of systems documentation (something returned to later on).

TABLE 5.1A
FREQUENCY OF USE OF FT PROFILE BY *GUARDIAN* JOURNALISTS:
ANALYSIS BY GENDER (Survey)

		Non-user[1]	Occasional user[2]	Regular user[3]	Heavy user[4]	Total
Male	n	22	17	16	14	69
	%	32	25	23	20	100
Female	n	9	5	2		18
	%	50	28	11	11	100
Total	n	31	22	18	16	87
	%	36	26	21	18	100

1. did not search themselves 2. search less than once a day
3. search once or twice a day 4. search more than twice a day

It appears that female journalists were less likely to conduct their own on-line searches than their male colleagues (Table 5.1A). Thus, while 21% of the respondents were women, they accounted for 29% of the non-users and just 13% of heavy users. Looking at the figures in another way, half of the female journalists who responded

categorised themselves as non-users (only a third of the men regarded themselves as such). There is no obvious explanation for this difference, but it is possible that: women reporters were more computer-phobic than men reporters, that women were happier to delegate their searching than their male colleagues or, maybe, women rated their abilities more honestly.

There were also very significant differences between departments in their use of PROFILE (Table 5.1B). In proportional terms Sport, from which, significantly, only 4 journalists out of 35 responded, had the highest proportion of non-users (75%) - sports journalists were also the library's lightest users. Part of the explanation for their low use lies with sports reporters legendary encyclopaedic minds, and part with the fact that they were well-served by numerous reference works. They were also more tabloid in tone and temperament than their *Guardian* colleagues and there might lie the principal reason for their low use of on-line and libraries. It might be a cultural thing as much as anything, for Vincent Hannah, a one-time political broadcaster, joined the Sports Department in 1995, has turned things on their head, by using PROFILE a lot. In terms of pure numbers it was in fact the Features department that has the largest number of non-users (13). This can be attributed to a number of factors: chief amongst them being that many Features staff were part-time reporters, with short-term contracts - they would have little time to familiarise themselves with the service. Another reason was alluded to earlier - they tend to be self-sufficient searchers, who only occasionally delegate the information seeking task: clearly this self-sufficiency extends to on-line databases, too. It must be said, however, that this does not necessarily mean that they could not benefit from on-line, because with the greater time they have available to search (not constrained as they are with daily deadlines) they would probably benefit more than the general news reporter - after all City journalists, who did detailed research were big on-line searchers as well. According to library staff the make-up of the department was changing, with many more young and inexperienced journalists entering the department - all resulting in an increased demand on the library and - as the logs will show - a decrease in end-use.

The Others category of journalists (composed largely of leader writers and senior editors) boasted the largest number and highest proportion of heavy users: 5 journalists, 42% of the department, searched PROFILE more than twice a day. Part of the reason for their high use lay with the fact that this group contained on-line (and computer) enthusiasts, who have been with the paper since PROFILE's introduction: some were party to the decision to install the system in the first place. There might also be an element of wanting to have been seen leading from the top in all this. If the

totals for heavy and regular users are combined, then City with 7 (77%) journalists, make a strong claim for the title of the heaviest on-line searching department. From the very outset City journalists have been strongly associated with information technology: they figured strongly amongst the early on-line leaders. Business data is of course very well catered for on PROFILE - and always has been, and here possibly lies a major explanation for its heavy use. Additionally, there was a stronger awareness of published information amongst City journalists, and also a tradition of helping themselves to it.

TABLE 5.1B
FREQUENCY OF USE OF FT PROFILE BY *GUARDIAN*
JOURNALISTS: ANALYSIS BY DEPARTMENT (Survey)

		Non-user	Occasional user	Regular user	Heavy user	Total
City	n	2	0	4	3	9
	%	22	0	44	33	100
Features	n	13	9	4	3	29
	%	45	31	14	10	100
Foreign	n	1	3	2	2	8
	%	13	38	25	25	100
Home	n	10	6	6	3	25
	%	40	24	24	12	100
Sport	n	3	1	0	0	4
	%	75	25	0	0	100
Others[1]	n	2	3	2	5	12
	%	17	25	17	42	100
Total	n	31	22	18	16	87
	%	6	26	21	18	100

1. includes journalists without a departmental allegiance. e.g. leader writers

Table 5.1C shows how end-use varied greatly among the different types of journalist - reporters, sub-editors, editors and specialists. Not surprisingly, perhaps, given their strong academic leanings, it was the specialists, who boasted the largest proportion of regular and heavy users (56% of them were of this type). In complete contrast, sub-editors - the poorest respondents to the survey, were also the lightest on-line users: very nearly three quarters (71%) did not search PROFILE, and there were no heavy users among them at all. Ironically, if anybody really needed to search PROFILE themselves, it was the sub-editors, for they frequently wanted information quickly and at a time when the library was unattended, and need the kind of simple fact checking that PROFILE excels at. In their case then, it is plain that need was not the sole determinant of whether they went on-line or not: lack of awareness and training were also important factors. Back in 1986, when PROFILE searching was in its infancy,

Guardian management actually thought that sub-editors would become PROFILE's most active users, but this has not turned out to be the case (Nicholas *et al*, 1987; 61).

TABLE 5.1C
FREQUENCY OF USE OF FT PROFILE BY *GUARDIAN* JOURNALISTS:
ANALYSIS BY OCCUPATIONAL CATEGORY (Survey)

		Non-user	Occasional user	Regular user	Heavy user	Total
Reporters	n	12	8	7	8	35
	%	34	23	20	23	100
Editors	n	3	7	3	2	15
	%	20	47	20	13	100
Sub-editors	n	10	3	1	0	14
	%	71	21	7	0	100
Specialists	n	6	4	7	6	23
	%	26	17	30	26	100
Total	n	31	22	18	16	87
	%	36	26	21	18	100

Computer logs

The logs provided contrasting, but complementary data on use, for what was being measured here was purely the number of times the journalists searched over a finite period of time - two days in this case. From the logs it was also possible to distinguish between the conduct of a **search session** and individual **searches** conducted as part of the sessions. A **search session** was signalled by a user logging on and off the system. A **session** might encompass more than a single **search**, users for instance might stack up their queries so that they can do them all at once. In the case of the intermediary it might have been expected that this was commonplace for they may be going on-line on behalf of a number of users. But as far as *Guardian* librarians were concerned it was not so much a case of stacking searches as being given new ones while they were on-line doing a search for someone else: this arises because of the sheer volume of searching and the speed with which it has to be accomplished. In practice it was not always possible to spot the difference because a user might be making a number of very different search statements to get at the answer they are looking for. Generally, though where new and conceptually different terms were introduced after a document display this was taken to be a sign that a new search was being undertaken. A single search may of course be conducted in more than one file.

According to the logs journalists conducted 86 search sessions, 99 individual searches over the two days monitored. Thirty nine journalists were identified as searching the system during that period, but the number was probably higher than that because there were seven on-line sessions that could not be assigned to anyone. It is estimated that around 170 journalists were working at the time of the computer monitoring, so approximately one-quarter of *Guardian* journalists went on-line during the two days. The questionnaire survey had indicated that 56 journalists were active on-line: the difference in the figures could be explained by the difference in date of the two investigations; by the difference in the news level at the time of the two studies (in newspaper terms there are quiet and heavy days) or by the fact that a number of journalist were not picked up by the computer logs because they were on leave (*The Guardian* is essentially a seven-day-a-week operation): the fact that the *proportion* of journalists going on-line was almost identical supports the latter explanation. Of the 39 searching journalists, most (77%) searched the system once or twice, although two journalists searched PROFILE seven times (Table 5.2). During the same two-day period 8 librarians at the newspaper conducted 53 search sessions and 69 individual searches on behalf of *Guardian* journalists. If the relative proportion of searches - journalists 59% and librarians 41% is compared to the relative proportions of expenditure, respectively, 70% and 30% it would appear that, outwardly anyway, the librarians offer better value for money. Although this indicates that *Guardian* journalists conducted more searches on PROFILE than the librarians, it has to be noted that during the two days the librarians also undertook 48 PROFILE sessions on behalf of *Observer* journalists, for whom they were also responsible. So, overall, the figures were quite similar. *Guardian* librarians believe that their figures are not a true reflection of the volume of their searching - they believe that an average of 100 searches a day for both papers would be closer to the mark.

TABLE 5.2
NUMBER OF FT PROFILE SEARCHES CONDUCTED
BY *GUARDIAN* JOURNALISTS (Logs)

No. of searches[1]	1	2	3	4	5	6	7	Total
No. of journalists	22	8	3	2	1	1	2	39
% of journalists	56.4	20.5	7.7	5.1	2.6	2.6	5.1	100

1. In two days

Of the identifiable on-line searching journalists 8 (20%) were women. This represented a slight improvement on the 1992 questionnaire, where 16% of users were women, though this is probably accounted for by an increase in women journalists

being employed by *The Guardian*. They conducted 21 sessions, 24% of all sessions. However, just 2 accounted for 62% of all on-line sessions conducted by women.

More Home journalists went on-line than any other group of journalists (10 in fact), but then Home is *The Guardian's* biggest department (Table 5.3). The questionnaire study also showed Home to have the largest number of regular or heavy users. However, if the calculation is done on the basis of the number of sessions conducted then Home were easily eclipsed by City - a much smaller department, which accounted for almost a third of all the on-line sessions conducted. Jacobson and Ullman (1989) also found business journalists to be the most active on-line users. Sport was the only department to produce no on-line users. This of course comes as no surprise for their low use of on-line was already well documented by the questionnaire study. Generally then, the computer logs concur with the survey data with the exception of the on-line searching strength of City journalists, which seems to have increased in the intervening years (1992-1995).

TABLE 5.3
NUMBER OF FT PROFILE SESSIONS CONDUCTED BY
***GUARDIAN* JOURNALISTS: ANALYSIS BY DEPARTMENT (Logs)**

	Number of users	% of all users	Number of sessions	% of sessions
City	8	20.5	28	32.6
Features	4	10.3	9	10.5
Foreign	6	15.4	14	16.2
Home	10	25.6	18	21.0
Sport	0	0	0	0
Others	6	15.4	12	13.9
Unknown	5	12.8	5	5.8
Total	39	100	86	100

N.B. Of all the sessions conducted, two produced no searches, and
for one of these sessions a journalist was identified.

5.4.1.3 Connect time

There were two related issues that required investigation here: (1) the total amount of time journalists spend on-line in a given time period; (2) the length of journalists on-line searches. While total connect time was a very good indicator of the amount of use journalists made of a system, the duration of individual searches was a sign of many things - level of expertise, type of searches conducted and searching style.

Survey

The questionnaire concerned itself with discovering the average time journalists' spent on-line during a search. This was more likely to be remembered than the actual amount of time they spent on PROFILE in the last week, month etc. Journalists did not keep records of their searches and nor did the software at the time provide a break-down for individual journalists. On the whole they claimed that they spent very little time on-line when they did a search. 78% of journalists said their searches averaged no more than 10 minutes in length, with only 4% admitting to conducting searches of more than 20 minutes in duration (Table 5.4). It was highly likely that journalists' time on-line was constrained by the short-time that they had available in which to conduct the search and the pressures for them to get off line so that their colleagues could go on-line (one password between a hundred or so journalists causes quite a bottleneck). It was also possible (but unlikely) that, thinking of cost constraints and what the library might think of them, they were not entirely truthful - and then there was the problem of recollection. However, the logs will show that the journalists were indeed very close to telling the truth.

TABLE 5.4
DURATION OF *GUARDIAN* JOURNALIST' FT PROFILE
SEARCHES: ANALYSIS BY DEPARTMENT (Survey)

	Minutes	1 - 10	11 - 20	20 +	Total
City	n		1	1	6
	%	67	17	17	100
Features	n	12	4		17
	%	71	24	6	100
Foreign	n	5	1	-	6
	%	83	17	-	100
Home	n	11	1	-	12
	%	92	8	-	100
Sport	n	-	1	-	1
	%	-	100	-	100
Others	n	8	1	-	9
	%	89	11	-	100
Total	n	40	9	2	51
	%	78	18	4	100

N.B. six journalists said their searches varied too much too say.

There were quite large departmental differences, with highly-pressurised Home journalists much more likely to conduct quick searches (92% of their searches were said to be over within 10 minutes), and Features and City journalists - two groups of journalist who were most likely to conduct on-line research - much more likely to take their time over searches (in both cases journalists claimed that 30% or more of their searches took longer than 10 minutes). The data provided by the survey, then, was very plausible.

Computer logs

The logs supplied both the total amount of time spent on-line by journalists and also the actual duration of their searches. Altogether journalists were on-line for just over eleven hours during the two day monitoring period. This meant that each end-user spent approximately 18 minutes on-line - just over 9 minutes per day: hardly figures to set management worrying. The logs provided confirmation that journalists' searches were very short indeed: well over half the search sessions and individual searches were over in five minutes or less, 4 sessions lasted just a minute (Table 5.5). Four minutes was most common session time, with a general clustering of times around the 3 - 6 minute mark. Thirty five minutes was the longest time spent on-line by journalists, but generally very few sessions lasted beyond 20 minutes - just 7% did so. There was a high degree of agreement as to what journalists said were their search times and what transpired in practice: 78% of journalists said their searches took no longer than 10 minutes and the logs wholly concurred with this, reporting that 79.7% of all searches were accomplished in this time. What is perhaps most surprising is that very short on-line times were achieved despite the displays of full-text records (see Table 5.19). Undoubtedly, what helped keep journalists searches short is the very limited range of terms and search options used (Tables 5.13 and 5.16), and the fact that they were in a far better position (than the intermediary) to make rapid relevance judgements.

, **TABLE 5.5**
DURATION OF FT PROFILE SESSIONS AND SEARCHES
AT *THE GUARDIAN* (Logs)

	Minutes	1 - 5	6 - 10	11 - 15	16 - 20	21 - 25	26 +	Total
JOURNALISTS								
Sessions	n	45	18	13	4	2	4	86
	%	52.3	20.9	15.1	4.7	2.3	4.7	100
Searches	n	57	22	11	4	2	3	99
	%	57.6	22.2	11.1	4.0	2.0	3.0	100
LIBRARIANS								
Sessions	n	23	14	7	5	2	2	53
	%	43.4	26.4	13.2	9.4	3.8	3.8	100
Searches	n	39	17	7	5	1	0	69
	%	56.5	24.6	10.1	7.2	1.4	0	100

5.4.1.4 Records displayed and screens viewed

Posting counts were not supplied with the computer logs, therefore other measures had to be turned to assess the volume of data retrieved by journalists. In fact two measures suggested themselves: (1) the number of records displayed; (2) the number of screens viewed. They are both imperfect measures though, and provide more of an

insight to the amount of data that is viewed as part of a search. Thus in the case of records displayed, patently this is not necessarily the same thing as the number of documents retrieved, although it could be. It is possible users retrieved more documents but chose to display only a proportion. In addition it was not always possible to establish the actual number of documents displayed, especially in the cases where users specified **all** in their display statement. Therefore the records displayed figures contained in Table 5.6 must be regarded as minimum figures. The number of screens the user viewed as a result of a search was another interesting but rough and ready measure. Again, this cannot be equated with the number of records displayed because the number of records displayed per screen is variable, being dependent on a number of factors: (1) the format chosen - the briefer the format the more records contained on a screen; (2) the length of the record - the longer the record the more screens needed to view it; (3) and, because it is a Carriage Return that signals a new screen, if users chose to override the natural break of 18 lines, with a **Nobreak** command then there was no knowing how many screens they viewed.

The number of records displayed in a search by a journalist varied enormously, from the 9 searches that displayed no records to the 2 searches that displayed more than 300 records (Table 5.6). While searches displaying between 1-10 records were quite common - 36% of searches fell into this category, it would be wrong to jump to the conclusion that journalists searches were characterised by the relatively small amount of documents retrieved: for, after all, the same table tells us that 30% of the searches saw the display of more than 50 records. A close inspection of the logs would suggest that the big differences in documents output can be explained by the different kinds of searches being conducted. Fact finding searches (biographical ones for instance) typically retrieve fewer items than the 'looking for ideas' type, which was characterised by wide sweeps of the database and much browsing of the retrieved records. However, even allowing for the under-reporting of journalists displays alluded to earlier, it does not appear that journalists' searches resulted in the retrieval of unduly large numbers of records. When it came to the number of screens displayed per search it was again the 1-10 band that featured most significantly: but this time accounting for an even higher proportion (46%) of all searches. There was very little clustering within this band - the whole numeric range being evenly represented. Generally, journalists viewed fewer screens than records, and this cannot simply be because more than one record can be displayed on the screen, because it is also possible for a single record to span a number of screens. Again, with three-quarters of all searches involving a display of no more than 20 screens, this does not appear to

constitute unusually high levels of document viewing. But neither does it constitute precision searching either.

The librarians registered virtually the same proportion of zero hits searches as the journalists - 17% to the journalists' 18%, but they differed markedly in two respects: (1) the proportion of searches which yielded 1-10 records - 20% in the case of the librarians, but 36% in the journalists' case; (2) the proportion of searches that yielded 40 or more records - in the case of the librarians the figure was 14%, whereas the comparable figure for journalists was 40%. So journalists' searches tended to polarise between the very low- and high-yielding searches, as was the case with politicians' searches of POLIS.

TABLE 5.6
NUMBER OF FT PROFILE RECORDS DISPLAYED AND SCREENS VIEWED
AT *THE GUARDIAN*: PER SEARCH (Logs)

			0	1-10	11-20	21-30	31-40	41-50	51-100	101-200	201 +
Journalists	Records	n	9	18	3	7	2	5	7	4	4
		%	18	36	6	14	4	10	14	8	8
Librarians	Records	n	6	7	7	8	2	1	2	1	1
		%	17	20	20	23	6	3	5	3	3
Journalists	Screens	n	9	42	26	4	3	3	4	0	0
		%	10	46	29	4	3	3	4	0	0
Librarians	Screens	n	6	29	11	9	2	1	0	0	0
		%	10	49	19	15	3	2	0	0	0

NB. It was not possible to establish records displayed or screens viewed for all searches because of the use of the use of the **all** command (in case of numbers of records displayed) and **nobreak** command (in the case of screen displays).

When it comes to screens viewed the differences between the two groups of on-line searchers were not so pronounced, although more journalists searches had screen displays of the order 11-20 screens and more librarians searches produced 21-30 screen displays.

5.4.2 Time of searching

It was once thought that, because journalists typically relayed their requests to the library to the last minute, that they would do the same with on-line databases. However, that does not appear to be the case, for Table 5.7 shows that journalists said that their searching was equally divided between mornings and afternoons. Only a little searching went on into the evening - and that was largely because most of tomorrow's stories would have been written by then. Features, which has more relaxed deadlines, were the main exception - 16% of their journalists searched after six o'clock. The computer logs (last two rows of Table 5.7) in fact confirmed that the

bulk of searching was equally distributed between the morning and early afternoon, but they also showed that searching was actually more spread out than the journalists' had indicated. This might be because increasing demands on the two passwords caused journalists to seek quieter times in which to search. The most recent conversations with the *Guardian* librarians suggest that, we might be witnessing a change in the way journalists work. Nowadays, for an increasingly large proportion of journalists, their first port of call is the library. This is to prepare them for the day ahead. At around 4 o'clock, another flood of journalists - sometimes the same people - come in and they are usually in a hurry.

TABLE 5.7
TIME OF DAY WHEN FT PROFILE WAS SEARCHED
AT *THE GUARDIAN* (Survey)

		10am-2pm	2pm-6pm	6pm-8pm	8pm-	Total
City	n	6	6	1	-	13
	%	46	46	8		100
Features	n	15	12	4	1	32
	%	47	38	13	3	100
Foreign	n	6	6	1	-	13
	%	46	46	8		100
Home	n	10	12	1	2	25
	%	40	48	4	8	100
Sport	n	-	1	-	-	1
	%	-	100	-	-	100
Others	n	9	9	-	1	19
	%	47	47	-	6	100
Total (survey)	n	46	46	7	4	103
	%	45	45	7	4	100
Total (logs)	n	36	35	13	15	99
	%	37	35	13	15	100

NB. The total exceed the number of journalists who searched because some journalists chose more than one time period.

5.4.3 File selection

At the time of the survey (1992) FT PROFILE hosted close to 50 individual files, and these files were also searchable in half a dozen or so pre-selected groups. So which files or file groups did *Guardian* journalists favour? According to Table 5.8 it was by some margin UKNEWS (a file group representing 28 UK national and regional newspapers - and half the files on the database). Of the files mentioned as being regularly used, UKNEWS accounted for half the mentions. After UKNEWS came *The Guardian* - cited 25% of the times. This illustrated quite clearly that journalists were largely using on-line to access external sources of information. Also, that they almost wholly relied on newspaper sources - not even wire services were used (in archival form).

The major features of departmental searching were: the fairly extensive use of *The Guardian* made by the Others - a category including journalists involved in the monitoring of the paper (however, even in this group a higher percentage preferred UKNEWS); Foreign's breadth of interest - demonstrated by their low use of *The Guardian* alone (accounting for only 12% the files mentioned), and their interest in the other papers (41%) and others category (29%), which included foreign sources. Indeed, every single department preferred access to a selection of papers rather than to just *The Guardian*.

TABLE 5.8
FT PROFILE FILES REGULARLY USED BY *GUARDIAN* JOURNALISTS
(Survey)

	Guardian		Other papers		UKNEWS		Others		Total[1]	
	n	%	n	%	n	%	n	%	n	%
City	2	15	2	15	6	46	3	23	13	100
Features	10	30	3	9	17	52	3	9	33	100
Foreign	2	12	7	41	3	18	5	29	17	100
Home	6	22	2	7	17	63	2	7	27	100
Sport	0	0	0	0	2	100	0	0	2	100
Others	7	41	2	12	8	47	0	0	17	100
Total	27	25	14	13	53	50	13	12	107	100

[1]Total times a file was mentioned. NB. Journalists chose more than one file, so base number exceeds the number of respondents.

Computer logs

In between searches PROFILE was parked for a few minutes in the file last selected by a journalist. As a consequence it was not always clear which file journalists were in or whether they actually wanted to be there. Thus the only occasions counted were where a journalist requested a file (using the **select** command) or where it appeared quite obvious (by the nature of the search) that they knew where they were. There were 83 occasions when it was felt that file selection was being made. On 53 (64%) occasions the selection made was UKNEWS, and on 11 (13%) of the occasions, *The Guardian* (Table 5.9). A further 14 files and file groups were searched, but not much.

The computer logs thus show a higher dependence on UKNEWS and a lower dependence on *The Guardian* than the survey indicated. This could well be because the UKNEWS file group has grown in number since the survey - and was thus more attractive - it does of course include *The Guardian*. More likely, the two days were insufficient to pick up the range of files that would have been used irregularly. In proportional terms the Librarians leaned more heavily on UKNEWS (67%) than the

journalists, but not quite so heavily on *The Guardian* (7.2%) - no doubt the proximity of the cuttings files were an important factor here. In addition, another 9 files or file groups were utilised.

Were journalists the single-source searchers that has been sometimes alleged? Well, not strictly, for the survey had shown that on average each journalist searched regularly about two files, and the logs - showing something a little different - recorded that files were changed in the cases of around 1 in 10 of searches. They were really not significantly more single-source oriented than the librarians, and, indeed, their searches actually ranged over a larger number of files, 10 of which were not used once by librarians. Librarians in turn used 5 files that were not used by the journalists. Whether the file selections made by journalists were apposite is quite another matter though, and one that cannot be answered here. The love affair both groups have with UKNEWS is surely due to its Siren-like qualities. However, it has to be understood that it can prove to be a tool for the lazy and unthinking: to search so many files, many of which, like *Lloyds List,* would be highly unlikely to be relevant is both time consuming (searches take much longer) and costly (the longer it takes the greater the cost).

TABLE 5.9
FT PROFILE FILES USED AT *THE GUARDIAN* (Logs)

File or File group	Journalists		Librarians	
	N	%	N	%
UKNEWS	53	63.9	46	66.7
Guardian (GDN)	11	13.3	5	7.2
Business Week (BW)	1	1.2	0	0
FINBIZ	1	1.2	1	1.4
FS	2	2.4	3	4.3
Financial Times (FT)	2	2.4	2	2.8
Hermes (HMS)	1	1.2	0	0
Independent (IND)	3	3.6	0	0
INTNEWS	0	0	3	4.3
McCarthy (MCC)	1	1.2	0	0
ML	1	1.2	0	0
Observer (OBS)	0	0	1	1.4
Observer/Independent	1	1.2	0	0
SCITECH	0	0	2	2.8
Standard (STS)	1	1.2	0	0
Tass newswire (TAS)	1	1.2	0	0
Telegraphs (TEL)	0	0	1	1.4
Times, Sunday Times (TIM)	2	2.4	0	0
TIM, TEL, GDN, ML	1	1.2	0	0
USNEWS	1	1.2	4	5.8
Washington Post	0	0	1	1.4
TOTAL	83	100	69	100

5.4.3.1 Checking hard copy resources before going on-line

A somewhat related issue to on-line source selection was source selection generally. One of the main reasons why so many librarians are sceptical about end-use is that they suspect that users, not only choose the wrong files in which to conduct their searches, but also go on-line without checking what is available in-house first. With the advent of on-line databases *Guardian* librarians have rationalised their cuttings system - selecting fewer cuttings overall and trying to create a complementary and integrated overall system by cutting those subjects less easily retrieved on-line: successful and cost-effective searching would require that both sources were consulted. As such, librarians regard it a waste of time and money for reporters to trawl through a database in search of information, which has already been paid for and is sitting on a (not so) nearby shelf, or available free at the end of a telephone line. Suspecting that journalists do not check the library's resources before they go on-line, they were asked that very question in the questionnaire. In response a quarter of respondents admitted that they *never* checked the cuttings first and nearly a half (46%) admitted to *sometimes* not checking the cuttings (Table 5.10). The possible explanation was laziness and a shortage of time, for the City Desk, who were amongst *The Guardian's* heaviest and most proficient on-line users, said they never searched on-line before checking hard-copy resources first. (Interestingly, one of INFROSS's main findings was that a characteristic of high volume information users was that they used all forms of publication heavily - Bath University; 1971). They were unusual, though, in having their cuttings files in the room in which most of them worked. In complete contrast, Home reporters, who had to climb a floor to their cuttings, and who were amongst *The Guardian's* poorest on-line searchers, not unsurprisingly, were averse to checking the files first - 50% of them said they never checked files before going on-line.

		Always	Sometimes	Rarely	Total
City	n	0	3	6	9
	%	0	38	62	100
Features	n	2	7	6	15
	%	13	47	40	100
Foreign	n	1	3	2	6
	%	17	50	33	100
Home	n	7	4	3	14
	%	50	29	21	100
Sport	n	1	1	0	2
	%	50	50	0	100
Others	n	2	8	1	11
	%	18	73	9	100
Total	n	13	26	18	57
	%	23	46	32	100

NB. Some journalist who did not search themselves answered this question (interpreting the question to mean before they asked the library to do a search for them, and also some journalists who did search failed to answer the question. Hence row totals do not match those of Table 5.1B

In an attempt to discover whether it was in fact ignorance (of what was available in the library), rather than access problems, that was the real explanation for poor consultation of in-house resources, the idea of putting a computerised index of cuttings files on the editorial system (ATEX) was floated to journalists. Most (77%) journalists thought in fact this would be a good idea. But in all probability, this would not alter the practices of the hard core of journalists, because 23% said they thought it wouldn't be useful - exactly the same proportion (and probably the same people) who said that they always searched without looking at cuttings. Even those who said that it would be useful admitted that they would be unlikely to check the index before starting the on-line search. Time is the essence in a newspaper and journalists are knee-jerk specialists when it comes to the use of information.

Guardian journalists were not alone in failing to check in-house resources before they went on-line, because this happens at News Internationals well (Erbach, 1994). The difference there though, was that library staff did not even expect them to check with the library first. News International library staff argued that journalists could not be expected to do this because: (1) library telephone lines were so busy that they could not always get through - and that they would clog them up even more if they attempted to do so; (2) journalists were so geographically removed from the library that they could not be expected to visit the library. Whether librarians always checked before going on-line is another matter.

5.4.4 Retrospective searching

Given journalists' preoccupation with events as they break and their stringent currency requirements, one might easily make the mistake of believing that they only ever searched back a few days or months when they went on-line. But this would be to misunderstand their use of on-line databases, for generally they use on-line to obtain background or contextual information to support stories they obtain from more current (the wires) and less public means of communication (the telephone).

Survey

In consequence the survey showed that over a fifth of journalists (21%) regularly searched back more than two years (Table 5.11). A number of journalists annotated their replies with comments like *as far back as she goes,* and *to the beginning of time* (in the case of the newspaper files on PROFILE that's typically 8-10 years). Indeed, less than a fifth of journalists (19%) said they searched back just six months.

Of the departments, Features, who were not generally quite as concerned as the other departments in highly topical events - and who tended to have more time to search - went back the furthest. One-third of their journalists searched back beyond two years. And, unexpectedly, it was not the City Desk - who dealt with very time-sensitive data, but Foreign, who proved most concerned with currency - nearly a third of their journalists conducted searches no further back in time than six months. This probably was because of Foreign's predomimant role in this country of updating current facts sent in by overseas reporters and planning/editing the foreign pages.

TABLE 5.11
RETROSPECTIVE SEARCHES OF FT PROFILE BY
GUARDIAN **JOURNALISTS (Survey)**

		1-6 months	6-12 months	12-24 months	24 + months	TOTAL
City	n	2	2	3	2	9
	%	22	22	34	22	100
Features	n	3	6	5	7	21
	%	14	29	24	3	100
Foreign	n	3	5	1	1	10
	%	30	50	10	10	100
Home	n	4	7	7	3	21
	%	19	3	33	14	100
Sport	n	1	0	1	1	3
	%	33	0	33	3	100
Others	n	1	2	6	2	11
	%	9	18	55	8	100
Total	n	14	22	23	16	75
	%	19	29	31	21	100

NB. Non-users answered this question, interpreting it to mean how far back would they wish to search

Computer logs

A search on PROFILE automatically searches back twelve months. By using the MORE command an additional twelve months will be searched. Repeating the command will take the search back to the beginning of the database: the actual date depending on when the file was first added to PROFILE, but in the case of most newspapers - the sources typically searched by journalists - to the mid to late 1980s. By counting the number of MORE commands used it was then possible to get a rough idea of how far back in time journalists searched. As the survey suggested journalists did search back quite a number of years: 52% of searches were conducted on the full retrospective file, 8% went back just 24 months. These figures showed that more retrospective searching was conducted than indicated by the questionnaire. The nature of journalism is probably changing, with more and more lifestyle pieces being produced - and these pieces are not as time-sensitive as news items are. By way of contrast, 67% of the librarians' searches were conducted on the full retrospective file, with another 24.6% going back just 24 months. The differences between the two groups could be accounted for by the librarians' drive for greater completeness in their searches, the fact that they are asked to do the longer searches, and the fact that increasingly more research is being handed over to them. It is also possible that journalists were not so conversant with the MORE command, though in fact you are prompted to use MORE after a result is posted on the screen. On the basis of both sets of data then hosts should be as concerned in keeping long-runs of newspapers on-line as in trying to ensure today's issue is available for searching tomorrow.

5.4.5 Types of search conducted

Journalists were given a choice in the questionnaire of what were regarded to be the main approaches that they took to PROFILE (Table 5.12). It was clear that when journalists searched FT PROFILE they, like many groups of end-users, were principally making subject searches and looking for background information . Three-quarters of searches were of this type In the searches of Foreign reporters (86%) and Home reporters (87%) this type of search was even more highly favoured. PROFILE searches do default to the subject search (no field indicators or codes are required) and this must reinforce the subject effect on a command-driven system, where the other search choices are not listed and, as a result, do not start with an equal choice of being picked. The moment you step outside of the subject search you need to know what the field operator is and what the field codes are. The computer logs throw more light on the nature of subject searching - much of the subject searching was of a biographical kind. Subject searching on PROFILE is essentially in natural language or free-text mode only - and this is thought to endear the system to journalists. One of the on-line-experienced journalists interviewed explained: *Free searching is what journalists are accustomed to, it's what they understand. As far as I can see, controlled vocabulary searching is always going to be a specialist preserve* [hence PROFILE's once studied avoidance of descriptors]. *It just isn't for the average journalist, whose focus is on something else - information dissemination, not information retrieval. Perhaps it's also a question of strategy: journalists are not strategic thinkers, they think sequentially.*

TABLE 5.12
MAIN TYPES OF SEARCH CONDUCTED ON FT PROFILE BY
***GUARDIAN* JOURNALISTS (Survey)**

		Date search	By-line search	Spell-check	Subject background search	Total[1]
City	n	2	1	2	7	14
	%	14	7	14	50	100
Features	n	5	0	4	21	31
	%	16	0	13	68	100
Foreign	n	0	0	0	6	7
	%	0	0	0	86	100
Home	n	2	1	0	20	23
	%	9	4	0	87	100
Sport	n	0	0	0	3	3
	%	0	0	0	100	100
Others	n	1	1	1	10	13
	%	8	8	8	77	100
Total	n	10	3	7	67	91
	%	11	3	8	74	100

[1]Total including others. NB. Some journalist mentioned more than one type.

A manual examination of the logs provides plenty of evidence that the vast majority of FT PROFILE searching concerned subject. Visual scans of the search terms used shows that as many as half of all subject searches involved individuals and organisations - name searches. The following list of terms, used in the first 20 searches conducted by journalists, gives a representative picture of what they were looking for and how they expressed it.

Fatty Arbuckle	*Robert Bolt*
Bruges, Football	*Intellectual Property Rights, Super Highway*
Auxerre, Football	*Pamela Banyard, Evan Dyer*
Gaidar, Soviet Union	*Lilly, Leeds*
Albright Wilson	*Kabul*
Mike Borlace, Mercenary	*Elisa Paris*
Legal aid, Fraud, Green form	*Sir John Bourn, Legal Aid*
Alan Travis, Prison, 150 million	*Gibraltar*
Hale Clinic	*Harry Ramsden*
Peter Pilkington	*Dorchester, Arabs*

From this list it can be seen that most of the terms and searches were highly specific. With the possible exception of the searches on Kabul and Gibraltar - and even they are not that broad in the light of English newspaper searches, there were none of the alleged unfocussed or broad searching here (*al la* Stanbridge, 1992). Where journalists got their width from was, of course, through searching a full-text database without specifying a field.

A clearer understanding of how journalists searched PROFILE was provided by the interviews - and in fact they showed that they used it much more creatively than at first appears. Take the example first of a City journalists - probably *The Guardian's* most proficient searchers. They used the system in four main ways:

(1) As a fact-checking facility - for example a journalist might want to know the relevant parties associated with a merger involving a stockbroking firm or simply check the publication date of a particular article or check what they wrote (reporters possess more information that is released through their stories and it is sometimes important to check what they have told their readers).

(2) As a research or investigative medium. For example, at a car company's annual meeting the chairman talked a lot about the Japanese car industry. There was nothing new or exciting about that, as he did that all the time - constantly

denigrating Japanese achievements in the field. However, with PROFILE the journalist can make a story about it. The number of times the chairman had repeated these criticisms could be checked and then an article written along the lines of *Chairman of company X attacks the car industry for the nth time . . .*

(3)   As a form of current awareness - one City journalist did his current awareness searches at lunch time. City journalists were probably unusual in running this type of search, though more recently (1995) foreign correspondents have been showing interest - the South African correspondent uses Compuserve's SDI facility. In 1986, when the researcher first put the idea of SDI to journalists it was received with interest, but unfortunately the databases at the time did not offer the required levels of currency: they do now, and one might expect to see SDI - once the domain of scientists only - being embraced by more (specialist) journalists. Journalists, of course, obtain much of their electronic current awareness from wire services, like Reuters.

(4)   As a means of gathering background or contextual data for a story or issue they were interested in. The following quote from a City journalist illustrates the kind of approach nicely: *There is a mass of detail passed on by companies over the years and you can only remember the main companies and personalities connected with them. Now there might appear on the tape* (of the wire service) *an item about a company that rings a bell; but you do not know in what context you've come across that company before. You would then go to World Reporter* [PROFILE], *look up the company name and suddenly the story is a story on Mr X rather than on a company nobody has heard of.*

Outside of the City department on-line was used largely as a producer of background data - in a very similar way to which journalists use their staple print-source, the cuttings file. A leader writer gave this example of a typical search: *Margaret Thatcher is much quoted as saying ' I have never said unemployment would fall'* [during her administration]. *Now this is the sort of comment that is easily remembered and misreported. When did she say it, and were those the exact words - PROFILE would provide me with the answer'.*

One columnist displayed another novel use of PROFILE, and that involved negative searching: *If I was concerned that I had missed some developments in a story that had dropped out of the limelight over the last couple of weeks, a search on PROFILE would confirm that nothing had been missed.* It would seem this type of search, together with its offshoot (*Has anyone written about this? No? Then I will*) - was often practised by journalists.

Of the remaining types of search, date searching was practised 11% of the times (and was most popular in Features, where it accounted for 16% of the searches), spell/style-checking 8% of the times and by-line (author) searches just 3% of the times. The PROFILE logs provided close corroboration of the figures for date searching: they showed 11% of searches employing the **pick date** command. Evidence obtained from the Table above would suggest that the reasons for the low use of the non-subject searches (date and by-line) might be partly due to their unawareness and the relative difficulty of executing these searches (field codes and indicators are not easily remembered). The logs were largely in harmony with all this: date and author featured in just 12% of all searches.

5.4.6 Search strategies and construction

The computer logs provided the opportunity to examine the search formulations and strategies of journalists to determine whether they were simple or complex in construction. The clues to establishing this were: (1) the number of terms used in a search; (2) the number of steps involved in the search; (3) the number of files searched. A fourth measure - the use of system commands - is considered in the next section.

Table 5.13 and the list of terms given above show that journalists' searches were economical in their use of terms. Single term searches were the most common (36.4% fell into this category) and two-thirds of all searches were expressed in two terms or less. The largest number of terms employed in a search was seven. Sullivan (1986) found that just over half of end-user searches contained between 2-5 terms, and in the case of this study the figure was not that dissimilar - 58%. On the whole the searches of the librarians employed more terms - only 27.5% of their searches depended on the use of a single term. At the other end of the scale 13% of librarians' searches employed 5 terms or more - the contrasting figure for the journalists was 9%, a significant, but not enormous difference. The explanation for the difference rests largely with the librarians greater use of synonyms and alternative word forms.

Terms[1]		1	2	3	4	5	6	7+
Journalists	n	36	29	18	7	3	4	2
	%	36.4	29.2	18.2	7.1	3.0	4.0	2.0
Librarians	n	19	18	18	5	4	2	3
	%	27.5	26.0	26.0	7.2	5.8	2.9	4.3

1. Number of distinct terms used in a search, including new terms introduced as part of a search on another file. Phrases were treated as single terms. Errors were not counted, but different word forms were.

On-line search strategies were largely direct and straightforward with 40% expressed just once in a single line, and over 70% expressed in two lines (Table 5.14). A number of searches were quite complex - or long-winded, involving 6, 7 or 8 steps but these were by far the exception. In the case of librarians the single search statement was even more marked: 45% were of this type. Generally speaking journalists' searches contained more steps, with over 14% of their searches involving 5 steps or more. The equivalent figure for the librarians was under 9%. Sullivan (1986) found that 30% of end user searches contained 3 or more steps - and this was exactly the same case with journalists (30%). The difficulty comes in knowing exactly what to read into this data. On the one hand searches involving a number of stages can be a sign of on-line maturity - the searcher interacting with the data and fine-tuning the search as a result. However, on the other hand, it could say something about the quality of the initial search expression. i.e. it was poorly constructed and thus had to be changed. In the case of the journalists, through ignorance of the searching procedures, they have more limited choices when they chop and change. What certainly inflated the step count for journalists was their preference for the get and pick incremental search procedure (see Table 5.16).

Even the librarians acknowledged that they sometimes go on-line without thinking, confident in their own skills: *half-way through we realise that we could do a better search and thus change our strategy - we could eliminate this waste by thinking before we go on-line.*(Presently there is a drive going on to ensure that staff map out their search statements before going on-line).

TABLE 5.14

NUMBER OF STEPS MADE IN SEARCHES OF FT PROFILE AT
***THE GUARDIAN* (Logs)**

Steps[1]		1	2	3	4	5	6	7	8+
Journalists	n	40	29	14	2	6	4	2	2
	%	40.4	29.2	14.1	2.0	6.1	4.0	2.0	2.0
Librarians	n	31	15	10	7	4	0	0	2
	%	44.9	21.7	14.5	10.1	5.8	0	0	2.9

1. This refers to the number of steps used in the search strategy - the number of lines used to specify the search. Reproducing the same strategy in another database was not counted here as an extra step.

Switching databases in a search can show a level of on-line maturity, though it could be argued that choosing the right one in the first place was the real sign of maturity. However, once on-line, journalist would appear to be essentially single source searchers - although of course that single source could be an 28 source file group, like UKNEWS. 10% of all their searches involved one file change, and 3% involved two changes (no journalist changed files more than that). The relative proportions for the librarians were identical. So on this measure the librarians do not exhibit any more on-line maturity than end users.

5.4.7 Use of system commands

Information professionals have long felt that end-users were not sophisticated searchers (e.g. Peters, 1989) and this was very much borne out by the results of this survey. Journalists maybe high-volume users, but they were also very limited in their use, knowledge and mastery of even the most basic on-line commands. Now FT PROFILE is far from being a sophisticated on-line service (hence its principal attraction to journalists) but even its relatively simple and small range of search devices were not fully employed or seemingly understood by *Guardian* journalists. It would appear that they relied almost wholly on the **get** (the principal search command) and **pick** (a 'masked' Boolean **and**) commands: most adopting a two-line **get** and **pick** approach to even the most simple of searches. A City journalist explained the common approach: *I do not search in a sophisticated manner, I use the get command and then refine if necessary using the pick command on date, period or topic. If I get fifty stories I am happy to trawl through the headlines* [the logs showed that over 30% of searches displayed that many, or more, documents]. For some journalists the searches were of an even more elementary level, thus a journalist searching for material on President Gorbachov, was observed entering the search **Get Soviet.** An answer in the region of several thousand documents was obtained, but this

did not worry the journalist concerned who proceeded to browse through the headlines.

Aware of the possible sensitiveness of journalists in regard to questions which might appear to portray them in a bad light, and also of the possibility of journalists saying that they used a command when they didn't, journalists were actually asked to reproduce the commands in the questionnaire (see Appendix 8, question 7).

The comma and the plus sign are significant symbols in PROFILE's search armoury: the former represents the Boolean **or** and the latter the Boolean **and** - you cannot enter the words **or** or **and** in full. The fact that, once connected to PROFILE, the plus sign on the ATEX editorial terminal operates on different - and not obvious keys - complicates its use somewhat. Table 5.15A shows that 67% of the journalists said that they did not use the plus sign in their searches, and that an even higher proportion did not use the comma (77%). What was interesting about this, was not simply the high levels of unawareness of elementary building block commands, but also the fact that, while the **pick** command is a viable alternative to the plus sign, there is no alternative to the comma - unless truncation is taken into account, and as will be seen that there is even greater unawareness about this. It is often asserted that end-users have difficulties with the concepts of Boolean logic (frequently confusing **and** and **or**), and maybe they have greater difficulty when they have to wrestle with symbolic equivalents - no matter how mnemonic (and PROFILE's symbols cannot be faulted on these grounds).

TABLE 5.15A
ON-LINE SEARCH COMMANDS USED
BY *GUARDIAN* JOURNALISTS (Survey)

		Brackets		Comma		Plus sign		Save		Access Search	
		yes	no	yes	no	yes	no	yes	no	yes	no
City	n	2	4	2	4	2	5	6	1	5	2
	%	33	67	33	67	29	71	86	14	71	29
Features	n	3	12	3	12	6	10	8	8	9	7
	%	20	80	20	80	38	62	50	50	56	44
Foreign	n	2	5	2	5	2	5	2	5	5	2
	%	29	71	29	71	29	71	29	71	71	29
Home	n	3	10	3	10	4	9	8	5	7	6
	%	23	77	23	77	31	69	62	38	54	46
Sport	n	-	1	-	.	-	1	1	-	1	-
	%	0	100	0	0	0	100	100	0	100	0
Others	n	2	8	2	8	3	5	8	2	8	2
	%	20	80	20	80	38	62	80	20	80	20
Total	n	12	40	12	40	17	35	33	21	35	19
	%	23	77	23	77	33	67	59	41	65	35

NB. 4 journalists failed to answer in the case of the first four columns; 3 in the case of the last column

Truncation - represented on PROFILE by an asterisk - was the least known of all the commands (Table 5.15B), though this finding was rather undermined by the poor framing of the question, which used the word truncation - a term that proved to be unfamiliar to journalists. But even so there was strong evidence to suggest that the concept was also not one familiar to journalists. 89% of journalists stated they were ignorant of the truncation command. Information professionals might justly wonder how journalists get over the vagaries of natural language indexing without resorting to the logical **or** or truncation. *Guardian* library staff were of the opinion that journalists just forget about apostrophes and different word-endings and assume the form of the word first thought of will do. The logs lend substance to this view.

TABLE 15B
FAMILIARITY OF *GUARDIAN* JOURNALISTS
WITH FT PROFILE COMMANDS (Survey)

		Byline		Date		Truncation	
		yes	no	yes	no	yes	no
City	n	2	5	3	4	1	6
	%	29	71	43	57	14	86
Features	n	6	9	6	9	2	3
	%	40	60	40	60	13	87
Foreign	n	-	7	3	4	-	7
	%	0	100	43	67	0	100
Home	n	2	11	1	12	1	12
	%	15	85	8	92	8	92
Sport	n	-	1	-	1	-	1
	%	0	100	0	100	0	100
Others	n	4	6	4	6	2	8
	%	40	60	40	60	20	80
Total	n	14	39	17	36	6	47
	%	36	64	32	68	11	89

NB. 3 journalists failed to answer this question

Now, the use of the **pick** command alone to associate two words or more in full text documents containing maybe several thousand words is highly likely to create many false associations (noise) and frustrations. It is only through field and word-proximity searching that this problem can be limited - limited is the operative word, because the problem can never really be overcome without some form of controlled language indexing - a form of indexing almost wholly absent from PROFILE. However, journalists were unable to (or couldn't remember how to) undertake an author (by-line) or date (dateline) search: 64% appeared ignorant of the former and 68% of the latter (Table 5.15B). The use of PROFILE's two word-proximity devices - / for relating words in a single sentence, and // associate words in the same paragraph -

was not questioned in the survey, but observation and interview at the time indicated that they were hardly ever used. Even seasoned end users appeared to be unaware of their obvious merits. Thus a frequent user of PROFILE, watching over the shoulder of a librarian as they conducted an on-line search on their behalf, showed considerable astonishment and interest when the librarian employed the / command. It has to be said that some good default programming on the part of PROFILE means that, for the commonest use of word-proximity (i.e. linking the words in a phrase together in the precise order input), journalists are not required to do anything other than enter the word string - stop words are automatically removed. PROFILE is on the whole very good at sheltering journalists from some of the rigours of on-line searching.

Perhaps, it is all a case of, if you don't know there is a solution to your problem, you don't worry about it? Observation would also suggest that some journalists, for some queries, do the association/refining of the data at the display stage - scrolling continuously on-line and skimming through large rolls of print-out when off line. But even here they do not always benefit from the systems full range of facilities. Take the following example of a journalist that has been using PROFILE heavily for some ten years: it was only this year (1995) when he awoke to the use of the **ctx** command, which enables located records (stories) to be displayed with the paragraphs containing the search terms only - primarily a keyword-in-context facility. Displaying lots of records in this manner is obviously quicker, more effective and cheaper than displaying the full-text of, typically 500 - 3,000 word documents, all the time - and without the help of a nearby printer. PROFILE does have automatic page breaks and this does limit the problem somewhat.

When considering on-line skills you cannot exclude ATEX commands from consideration, because as far as users are concerned its all very much part of the same search process. In recognition of this journalists were questioned about an important ATEX feature: its automatic save facility - by which the search session was saved until the next one was commenced. This was done in recognition of reporters' natural inclination to squirrel away data for further processing at a later time and also to compensate for the remoteness of the printers. Given the very practical nature of these concerns it was thus not surprising that a good number of journalists used the save and accessing process associated with the operation: 59% used the save facility and and 65% knew how to access it afterwards (there must have been some misunderstanding here because, logically, it should have been the other way round).

On-line skills (or their absence) were not evenly spread amongst *Guardian* departments. Thus City came out as the most proficient department: 86% of its journalists said they used the save facility and 71% knew how to access the saved search afterwards; 33% used the comma (the Boolean **or**) in their searches and 29% used the plus sign and a very creditable 33% said they employed brackets to nest their search terms. City journalists were also the most knowledgeable with, 71% being able to construct a byline search, 43% being able to execute the date search, and 14% being able to handle truncation. Home appeared to be the least proficient and aware department, with three of the eight commands listed obtaining a 15% or less awareness rate. It is surely no coincidence that City were the paper's busiest and most self-sufficient searchers and Home the lightest and least self-sufficient. Regular use obviously goes hand and glove with proficiency.

Interviews with journalists highlighted a problem of on-line searching not so far touched upon - the problem of using hyphenated words. The problem arises because they were unaware that the hyphen has a special meaning and function on PROFILE - it is the system's **not** symbol: hyphens embedded in phrases are treated as spaces. This can lead to confusion as the following search by a journalist illustrates. Information on the non-proliferation of arms treaty was required. The journalist keyed in **Get non-proliferation treaty.** In return he obtained 5,984 postings in which **non** occurred, but **proliferation treaty** did not. After the penny had dropped, he laughed about it.

Computer logs

Through the computer logs it was possible to see whether the survey data tallied or whether there had been any changes in the use of commands. It was also possible to monitor the use of many more commands than journalists could feasibly be questioned about. Indeed, the use of 19 of PROFILE's search commands were examined - even a simple system has that many. **Get** was by far the most used command (used 169 times) but there were no surprises there as all searches have to commence with the command (Table 5.16). With 99 searches having been conducted altogether, each search averaged nearly two **gets**. Because with **get** you are starting a search anew this meant that users were making on average two stabs at their searches. On average 71% of searches contained a **pick**; and 41% of all individual **get** searches were followed up by a **pick.** Pick was the second most popular command, used on 71 occasions. The Table confirms the classic **get** and **pick** mix of most end-user searches noted by the survey and Miles (1993). The only other command used with any frequency was the Boolean **and**: two-thirds of searches featured **and**. Journalists were obviously alert to

the difficulties of using the ATEX plus sign. Given that the **pick** is in fact a second-step Boolean **and,** it can be said that most searches were constructed with an **and**. In a full-text system with typically 500-1,000 word documents the **and** does not have a lot of potency, certainly not the potency it has in a bibliographic system. The more effective method of relating terms through word proximity was used only occasionally - just 8% of all searches used sentence proximity, and none paragraph proximity. In fact journalists achieved a degree of precision (knowingly or not through the phrase search - a search on PROFILE which can be done without the aid of any command or device: a total of 101 were employed in 99 searches and 43% of the terms contained in either a **get** or **pick** statement were phrases. Harold Jackson, *The Guardian's* system editor was quoted earlier as saying that four commands was about all journalists could manage - they seem to get by with three, perhaps, the fourth is a display command?

The fact that the Boolean **or** was not used at all and truncation just once indicates that journalists hardly ever present the system with a choice of related or synonymous terms - and searching can only really be conducted through natural language terms. The third Boolean operator **not** was also wholly neglected. Journalists did not even use it in error conducted in error. In the light of what has been said it is not surprising that the bracket - the means by which **or** and **and** logic can be combined in a single search statement, was never used -although a small proportion of journalists in the survey said that they used it. The concept of brackets altering the priority of processing the search statement is not an easy one to grasp. Indeed, if you had to choose one feature that really differentiated end-user searching from that of professional librarians, the brackets would surely be it.

TABLE 5.16
USE OF FT PROFILE SEARCH COMMANDS AT *THE GUARDIAN*
(Logs)

COMMAND	NUMBER OF TIMES USED IN SEARCHES		AS A % OF ALL SEARCHES UNDERTAKEN		AS A % OF ALL GETS USED		AS A % OF ALL GETS AND PICKS USED	
	JNLSTS	LIBS	JNLSTS	LIBS	JNLSTS	LIBS	JNLSTS	LIBS
GET	169	124						
PICK	70	35	70.7	50.1	41.4	28.2		
TRUNCATION	1	40	1.0	58.0	0.6	32.2	0.4	25.2
AND	36	22	36.3	31.2	21.3	17.7	15.1	13.8
OR	2	26	2.0	37.7	1.1	21.0	0.8	16.4
NOT	0	0	0	0	0	0	0	0
BRACKETS	0	11	0	15.9	0	8.9	0	6.9
PHRASE	101	82	102.0	118.8	59.8	66.1	42.5	51.6
SAME SENTENCE	8	28	8.1	40.6	4.7	22.6	3.3	17.6
SAME PARA.	0	19	0	27.5	0	15.3	0	11.9
START	0	19	0	29.2	0	15.3		11.9
TITLE	4	9	4.0	13.0	2.4	7.3	1.7	5.7
AUTHOR	1	7	1.0	10.1	0.6	5.6	0.4	4.4
GET DATE	0	1	0	1.4	0	0.8	0	0.6
PICK DATE	11	1	11.1	1.4	6.5	0.8		
SAVE/RUN	0	13	0	18.8	0	10.5	0	8.2
STEPBACK	3	7	3.0	10.1	1.8	5.6	1.3	4.4
REVIEW	0	1	0	1.4	0	0.8	0	0.6
HIGHLIGHT	0	1	0	1.4	0	0.8	0	0.6
TOTAL	406	446						

There are various ways this table can be read. The first column basically gives raw figures of command usage. Column 2 relates command usage to the number of searches conducted, whereas column 3 expresses command usage as a percentage of the individual main (get) searches and the final column (4) shows command use as a proportion of both main and secondary (pick) searches.

PROFILE has a very limited range of fields - especially when compared with POLIS. Even so field searching was hardly ever practised: title (headline), and author (byline) were on average employed in respectively 4% and 1% of searches, and no title and leading two paragraphs (start) searches were conducted at all. Date searching obtains more prominence on PROFILE - it has its own get and pick commands - and in fact proved more popular, with 11% of searches specifying a date. The survey data showed that around a third of journalists said they used author and date searches but obviously not that regularly.

40% (8) of the commands listed in Table 5.16 were not employed at all, indicating that end-users used a limited range of commands: these commands variously offer increased precision (**not** for instance), control of the search process (**review** for instance), speed (**save/run**) and convenience (**stepback**). The main uncertainty in explaining away the low use of PROFILE commands is whether to attribute it largely to the broad brush information seeking styles of journalists, to inadequate training or to the fact that, as one senior journalist put it, *they fundamentally don't grasp the very nature of the database - this is a much deeper problem than the blockage of the hardware* (Nicholas, D *et. al*, 1987). Librarians though are not short of explanations: *... what is universal is that they all use PROFILE very badly. You watch them sometimes* [when] *they come up here* [to the library] *to search if the line is busy downstairs. We're horrified when we watch. We can't believe what they're doing.*

Their skills are very basic. Basic searches they may be, but not necessarily bad searches.

As might have been expected the librarians employed a much fuller range of command, using all the commands, bar one (the confusing **not** sign). The librarians made far greater allowance for word variants and synonyms thus on average 40% of their searches featured truncation and nearly 38% the Boolean or: the corresponding figures for journalists were 1% and 2%. Very, very big differences here. They also showed a much stronger preference for word proximity searching - no journalist employed the *in the same paragraph* feature, but nearly 28% of the librarians did, and proportionally five times as many of their searches used the *in the same sentence* word proximity facility. Further confirmation of the use of word proximity by librarians was provided by a one-day survey (sample size n=77), conducted by library staff into their use of PROFILE commands. This showed that 47% of their searches included word-proximity devices. Field searches were also a favourite of the librarians, especially the *title and the first two paragraphs* (**@ START**) search - one that attempts to circumvent the problems of vague or ambiguous newspaper title searches: 29% of the librarians', searches featured the command. (The self-same day survey put the figure at 25%). The **@ START** command was especially used by librarians for biographical searches, full-text searches created too much noise, particularly for someone like John Major, who might get a mention a hundred times a day in the database, burying the searcher in a sea of documents.

Not only did librarians use a wider range of commands - and this is obviously related - they also used more commands per search. Thus the 99 searches conducted by journalists generated 406 search commands - an average of 4.1 per search, whereas the librarians 69 searches generated 446 commands - an average of 6.5 per search.

Errors
The computer logs offered the opportunity of examining journalists' searches for errors in construction and execution and in so doing, possibly, corroborate some of the librarian's anecdotes. It was not always easy to determine whether something was an input error or a faulty command. There was also the problem of whether a search was so broad or loose as to constitute a logical error - in practice this proved too difficult to determine in the light of journalists searches and the benefit of the doubt was given to the searcher. However, journalists' searches were not so much characterised by misuse of the search and display commands - there were 20 instances when misuse was detected, as by typographical and spelling errors - these were

detected on 44 occasions. Indeed searches were generally so simple in design and execution that there was probably very little room for error in the use of commands. As expected librarians made far fewer errors with the commands (6: on average 9% of their searches contained errors compared to the journalists' 20%), but they did make more spelling and typographical errors: 45, just one more than the journalists, but proportionally 65% of their searches contained errors and only 44% of those of journalists did so. The fact that, when searching PROFILE on the ATEX terminals, you cannot backspace must be partly responsible for some of these errors. Also the speed with which they search must take some of the blame.

5.4.8 Displaying/viewing records

PROFILE is a full-text host and as a result it functions both as a library catalogue and as the library's shelves. And, to take the analogy further, all the walking and browsing is done at the terminal through a range of display commands. As a consequence the utilisation and inter-play between these commands is very important in a full-text system. Consider the range and function of PROFILE's display commands. Firstly, there are the commands that determine the amount and nature of detail displayed: **h** provides the title and, in some cases, the section heading; **ctx** provides the title and paragraphs containing the highlighted search terms (this form of display is highly variable in length - two lines to several screens - and maybe all that is needed to find the required information); **tx** provides the full-text of the document. Accompanying these commands would be a number indicating how many documents were to be reviewed, or, in the case of a display of everything in the set, the word **all.** In a full-text system so much data is scrolling around the screen that PROFILE provides an automatic page break after 18 lines - normally a screen's worth of data. This generally works fine until you want to display a number of full-text documents that you have already determined are relevant - in this case having to respond to the system message after ever screen can be very tiresome. In recognition of this PROFILE enables you to cancel the break, with the command - **nobreak.** The use of the **nobreak** command can prove disastrously expensive, with a system that charges by the line displayed. If you make a mistake, you will watch screens of expensive information scroll by, and by.

The order in which display commands are used and their combination depends partly upon the number of hits obtained, but a logical approach would be to: view records first in title format, scanning them for relevance; then, with the screen break off, display obviously relevant ones in full-text and then doubtful ones first in context and

then in full-text, if they turn out to be relevant. Although the whole process is not unduly complicated it does require a degree of thought and the display of a certain amount of on-line dexterity. Also, while you can get by without knowing many of the search commands, it is a different story with the display commands. Thus it was no surprise to find that journalists were generally more proficient in displaying data than they were searching for it, but then there were far fewer command to master.

Despite their very short searches journalists managed over four display commands per search (Table 5.17) - the same as the number of search commands used per search. A visual scan of the logs confirmed that journalists carefully go through records on-line to establish relevance. As indicated earlier the approach preferred was the two-prong **headline** and **text** display. The **context** display was quite often by-passed - the full text display being used more than twice as much. There are three possible explanations. Firstly, journalists were simply not aware of it. Secondly, they felt that they just did not need it - the number of records retrieved were sufficiently manageable not to warrant its use (a high proportion - 54% - of journalists searches retrieved 10 records or less). But in all probability it was the fact that so much journalism is looking for small detail - take the earlier example of the Thatcher quote (page 191), and a **ctx** display of 10 items is obviously quicker and easier than wading through stories of 500 words or so. Thirdly, and most likely, journalists were deeply suspicious of partial data. This suspicion is based, in part at least, on what appears to be a complete misunderstanding of databases in general and PROFILE in particular. A *Guardian* librarian provided this example in illustration. *Say you are looking for something on Arizona and you find a story whose headline clearly indicates that it is an article on Lancashire. In this circumstance we would give the journalist a **ctx**. But I have had people insist, despite my explanations and protestations, that I print out the full text. On having done so, they say, in disappointed tones ' oh but that's the **only** mention of Arizona in the article'.* In passing journalist are even more suspicious of abstracts - and on this they are on better ground (see Nicholas *et al.*, 1987: 99).

TABLE 5.17
FT PROFILE DISPLAY COMMANDS USED AT *THE GUARDIAN*
(Logs)

Display command	Journalists			Librarians		
	n	%	Ave. per srch	n	%	Ave. per srch
Headline (H)	191	45.0	1.9	102	35.2	1.5
Context (Ctx)	69	16.3	0.7	110	38.1	1.6
Text (Tx)	164	38.7	1.7	77	26.6	1.1
Total display cmnds	424	100	4.3	289	100	4.2
Nobreak (nbrk)	13	-	0.1	49		0.8

Librarians used an almost identical number of display statements per search (4.2), so both groups would appear to view the data with the same frequency, but not in the same manner however, for the librarians - possibly, with cost in mind - opted for more context displays (22% more) and less full text ones (12% less). They were also more conscious of the cumbersome and unwieldy nature of full text displays (see following paragraph). Librarians also used the headline display less frequently, using the context display in preference, as they were probably all too aware of the vague and misleading nature of newspaper headlines. The one-day survey mentioned earlier also pointed to a strong use of the **ctx** display, being used in nearly 80% of the librarians' searches. In discussions with the librarians they argued that **ctx** is the most important command on PROFILE, and they would not subscribe to a system without a similar facility. Patently, display commands take on an enormous amount of importance in the case of full-text systems.

Mention has been made earlier of the tedium associated with full text displays and how the use of the **nobreak** command can alleviate some of that tedium. Clearly librarians were more conscious of the problem (and aware of the command), for they used it many more times than the journalists - 49 times as against 13 times for the journalists. Even so the fact that it was used 13 times meant that it was far more used than most search commands - only four of these were used more frequently than that. It is also not an easy command to remember - it is so easy to enter no break, nobreaks or no breaks. In the case of the librarians it was used principally for searches which would be sent on to journalists via the editorial system (ATEX).

5.5 User satisfaction and problems experienced

Generally then, it has been established that journalists were relatively high-volume on-line users, who tended to search simply. But how successful were they as searchers? After all mastery of systems commands alone does not guarantee success in on-line searching, in the particular case of PROFILE having a good natural language search vocabulary is also extremely important. Given that journalists are word smiths by profession it might be expected that they would be good at choosing productive keywords. Also, observation - and some of the survey answers - would lead the professional onlooker to the view that journalists' understanding of the on-line search process was poor - or at best limited, and this necessarily left them dissatisfied with their endeavours. In fact it was not quite as simple as that for they seemed generally pleased with the fruits of their on-line labour: well over a quarter (28%) said they were nearly always successful with their searches and another 53% were sometimes

completely successful (Table 5.18A). And this despite the fact the question was phrased in such a way as to discourage journalists from opting too easily, or unthinkingly, for high levels of satisfaction. Of course, a cynic might argue that these levels of success were artificially high because journalists did not know what they missed and, if given this information, they would soon change their views on how successful their searches were - but this applies to the professional searcher too - and they rarely know what they missed. It is always possible that, these relatively high levels of success obtained, because successful searching of full text natural language retrieval systems almost always entails a lot of browsing through large amounts of data - something which journalists (and politicians) have a lot of experience in. Furthermore, because journalists search on a whim or hunch, they are in a far better position to make relevance judgements than the librarians. Media librarians are also very news conscious too, and when they are delegated a piece of research in its entirety, they may search on a whim or hunch as well. For what is good for journalists will often be good for the librarians - they are after all searching on behalf of journalists.

TABLE 5.18A
SUCCESS OF *GUARDIAN* JOURNALISTS' FT PROFILE
SEARCHES: ANALYSIS BY DEPARTMENT (Survey)

		Nearly always	Sometimes	Rarely	Total
City	n	1	5	2	8
	%	13	63	26	100
Features	n	4	9	5	18
	%	22	50	28	100
Foreign	n	2	5	0	7
	%	29	71	0	100
Home	n	4	7	3	14
	%	29	50	21	100
Sport	n	0	0	1	1
	%	0	0	100	100
Others	n	5	5	0	10
	%	50	50	0	100
Total	n	16	31	11	58
	%	28	53	19	100

NB. Two journalists from Features and one from City answered this question in the
light of their previous experience as end-users - they were no longer active searchers.
In the case of Home, one journalist did not answer the question.

Not everybody experienced success or satisfaction - and it is so easy to play up the successful figure, so who amongst the journalists were then the most/least satisfied with their searches? Of the departments the Others were clearly the most satisfied - 50% said they were nearly always successful with their searches, and Features the

least satisfied - 28% said they were rarely successful. Interestingly, Home Affairs journalists who scored very low on their use and knowledge of PROFILE commands (see Table 5.15A/B) had the second highest proportion of journalists who said they were nearly always satisfied with their searches. On this evidence lack of systems expertise is no barrier to high perceptions of success. Of the occupational groups, reporters - a large and important group, were the most unsuccessful, with only a quarter (26%) saying that they were nearly always successful with their searches, and the Editors the most successful - half saying that they were almost always satisfied with their searches (Table 5.18B). Female journalists, who as we have pointed out, made less use of the system than male journalists, were also more likely to rate their searches as more unsuccessful than men: 45% of the women said their searches were rarely successful, but only l3% of the men thought so (Table 5.18C). Perhaps this fact partly explained their low levels of on-line use reported previously.

TABLE 5.15B
SUCCESS OF *GUARDIAN* JOURNALISTS' FT PROFILE SEARCHES:
ANALYSIS BY OCCUPATIONAL CATEGORY (Survey)

		Nearly always	Sometimes	Rarely
Reporter	n	4	16	7
	%	15	59	26
Editor	n	5	4	1
	%	50	40	10
Sub-editor	n	2	2	1
	%	40	40	20
Specialist	n	5	9	2
	%	31	56	12
Total	n	16	31	11
	%	28	53	19

TABLE 5.18C
SUCCESS OF *GUARDIAN* JOURNALISTS' FT PROFILE
SEARCHES: ANALYSIS BY GENDER (Survey)

		Nearly always	Sometimes	Rarely
Men	n	14	27	6
	%	30	57	13
Women	n	2	4	9
	%	18	36	45
Total	n	16	31	11
	%	28	53	19

Logically, it might have thought that the heavy on-line users would be more successful with their searches. But this was not really borne out by the data (Table 5.18D), for while a marginally higher proportion of heavy users were in fact nearly always successful with their searches, this only amounted to 1 percentage point higher than the that for occasional users and fewer regular users were rarely successful with their searches. There are probably two explanations for this: firstly, high levels of use do not necessarily guarantee success in searching, although it does seem to increase awareness; secondly, the high volume, proficient searchers were that more critical of their searching and that much more aware of their frailties.

TABLE 5.18D
SUCCESS OF *GUARDIAN* JOURNALISTS' FT PROFILE SEARCHES
ANALYSIS BY TYPE OF ON-LINE USER
(Survey)

		Nearly always	Sometimes	Rarely
Occasional user	n	6	8	6
	%	30	40	30
Regular user	n	5	15	2
	%	23	68	9
Heavy user	n	5	8	3
	%	31	50	19
Total	n	16	31	11
	%	28	53	19

It is possible to explain journalist searching in another (more positive) light, and by so doing, square the high levels of satisfaction with the low levels of system knowledge. Firstly, journalists, through their copy tasting routines, are well used to skimming through large volumes of data and making rapid relevance judgements - there is nothing new in wading through data on-line to make similar high speed relevance checks. Secondly, so-called simple searching, provides them with high recall, which in turn provides a suitable framework for browsing for their ill-defined queries: the refining/honing of searches was done visually - an over-refined search might not provide sufficient noise to feed off. Thirdly it was simply just a case of what they didn't know about they didn't worry about: and journalists were rarely put in a position of seeing what a well-structured search would produce. Lastly, given the very tight deadlines journalists had to work with and the rich array of information sources available to them, they were quite happy to grab at information on-line: it was not the end-all or be-all.

While the published literature is full of end-user satisfaction surveys, it is hard to find one that asks librarians how successful they though their searches were - maybe

because there is a presumption that most librarians' searches were successful - their greater use of system commands and higher levels of training would guarantee that. Indeed this seems to be very much the case because when *Guardian* librarians were asked the same question as that asked of the journalists, seven (86%) said they were mostly nearly always successful with their searches. Much, much higher levels of satisfaction. The remaining librarian felt that success varied enormously.

Most of the librarians did hedge their answers though, saying that it was difficult to know for certain in the absence of feedback from journalists. They did not get much in the way of that: *Many journalists never reply assuming that what you have sent them is what is there.* Some pointed to the fact that they had more success with particular enquiries than others. The more factual, the more specific and the better articulated the enquiry, the more likelihood of success. *Woollier enquiries, like 'give me stories on famous jealous lovers', where there is no real correct answer are more difficult to search for. When the journalist's request is more vague you need to use you own initiative in weeding out relevant information it is more difficult to judge.* One librarian provided one reason why journalist searches were not quite so successful: its *good to avoid their helpful hints about what terms you should search under.* Maybe being a word smith does not help you to search on-line?

The problems journalists experienced when searching came out in their reasons for delegating the search. What came out as surprising was that the principal difficulty was not that of access - and you would have expected that as a result of 100+ journalists jockeying for one of the two passwords, but of framing the search query. This is dealt with in more detail in section 5.7.

Computer logs

Of course, computer logs give no hard evidence of how successful or satisfied journalists were with their searches, but the logs did provide some clues. Firstly, there was the records displayed data. Something positive can be read into the fact that records were actually displayed as a result of a search. However, this is not a strong measure of satisfaction for allowances have to be made for the fact that negative searches can be positive, and that a display of records does not necessarily mean that useful records were found. Secondly, there was the data on whether the search ended with a full text display? This was the stronger measure of the two, because a display of headlines might only provide confirmation that the search was poor, whereas a full-text display usually follows a context or headline relevance check - presumably then, the full-text displays mean relevant items were found and a degree of success was

achieved. There is an analogy here with books located and then borrowed in a library - there is a presumption of relevance in both cases. The measure is itself suspect on the grounds that it is possible for journalists to obtain the data they wanted from a headline or paragraph display, or, indeed, from a single record. The actual number of documents displayed (Table 5.7) can also provide some evidence of success, but in the particular case of journalists this measure loses some of its potency for scrolling through screen after screen of document displays can be very much part of the journalist's searching style. The number of documents displayed can probably be best used as an activity measure.

Using the first possible measure of satisfaction or success - document displays, journalists must be reasonably happy with their searching for only 18% of their searches resulted in no displays of documents (Table 5.19). Interestingly, this figure almost exactly corresponds with the questionnaire response on satisfaction - 19% said that they were rarely satisfied. Of course we are talking about two slightly different things: the first tells us what proportion of *searches* were successful, whereas the second tells us the proportion of *journalists* that were successful. According to the second measure - full-text displays, journalists were again reasonably successful, but not quite so successful - 63% of all searches resulted in the retrieval of full text-documents. On that score then journalists obtained a two-thirds success or satisfaction rate. For the librarians the no display figure was almost the same (17%), and the full text figure somewhat lower (60%), but then librarians might sometimes have felt that the context display provided all the information that was required. On the whole though the logs and the questionnaire responses were in broad agreement about levels of success and satisfaction. There was not much evidence in Table 5.19 to indicate that journalists wallowed in the data generated by a search: the average search generated between 1 and five documents, though in comparison to the librarians - and this is where the rumours possibly come from - they did conduct a larger proportion of searches which produced full-text documents.

TABLE 5.19
EVIDENCE OF SEARCH SUCCESS AT THE GUARDIAN:
FULL TEXT RECORDS RETRIEVED PER SEARCH (Logs)

	Full-text displays	0	1 - 5	6 - 10	11 - 15	16 +
Journalists	n	37	46	5	9	2
	%	37.3	46.5	5.1	9.1	2.0
Librarians	n	28	28	9	2	2
	%	40.1	40.1	13.0	2.9	2.9

5.6 Training

One of the arguments against allowing end-users to search expensive databases themselves is that they are insufficiently trained and are, anyway, ambivalent about maintaining their skills. As a consequence they represent a costly liability to any organisation that is 'rash' enough to provide for end-use. In the past *Guardian* management has been able to duck the issue by opting for unlimited-use agreements, for in such circumstances they paid the same no matter what the quality of searching was like. So why engage in costly and disruptive training programmes, for which most journalists would not even bother to turn up? It may be hard to believe (although their searching characteristics do bare witness), but hardly any *Guardian* journalists have been trained on PROFILE - some limited training was provided by the Host in 1986, but of the few that were trained few remain. Even at News International, where, if anything, end users were thicker on the ground, few journalists have received training - 1986 being the last time newspaper-wide training was tried. With the ending of unlimited use agreements by PROFILE all that has changed and untrained end-use now brings with it severe cost penalties. Even at the *Financial Times*, where training generally obtains a higher profile, it is more than five years since journalists obtained any training (Batten, 1994).

Believing that the limited range of commands used by journalists might be a function of lack of training, *Guardian* journalists were asked if they would be willing to be trained in the use of PROFILE. There certainly was nothing ambivalent about their response, for more than 90% said that they would like training, and both non-users and heavy users were of the same opinion (Table 5.20). Indeed, there was a certain desperation and urgency about many of the replies - some pleading to be provided with 'child's guides'. Notably, the heaviest and most proficient journalists (City for instance) were every bit as keen - indeed, sometimes more keen - as the lightest and least proficient of the journalists (Home).

However, despite the best efforts of the Library staff, few journalists came forward for training, making one believe that when it comes to the crunch journalists' prefer to spend their times otherwise engaged. There did not appear to be a willingness to invest time and effort, and to get to grips with the finer points of searching or source selection. Library staff were of the opinion that unless *Guardian* management made training official and compulsory and programmed it into journalists' working schedules, things were unlikely to change. However, even this strategy would be probably doomed to failure, for News International - in response to journalist' demand - tried organising training through the editors, spreading training sessions

liberally through the day, and yet only a handful of the several hundred invited journalists turned up. (Erbach, 1994).

Anyway, management were probably unlikely ever to grasp the nettle, because the issue of training is a two-edged sword, for while proficient users would mean better searches, less waste and would free the system for others more quickly, it would also, surely, lead to more use and increased cost.

TABLE 5.20
GUARDIAN JOURNALIST'S WILLINGNESS TO BE
TRAINED ON FT PROFILE (Survey)

		Willing	Not willing	Total
City	n	9	0	9
	%	100	0	100
Features	n	26	1	27
	%	96	4	100
Foreign	n	7	0	7
	%	100	0	100
Home	n	20	5	25
	%	80	20	100
Sport	n	3	0	3
	%	100	0	100
Others	n	10	1	11
	%	91	9	100
Total	n	75	7	82
	%	91	9	100

NB. This question was answered by some non-users as well.

5.7 Delegation of the search

So far we have largely being discussing the on-line users, but it was clear that not all journalists at *The Guardian* wanted to search FT PROFILE themselves. In fact, the questionnaire indicated that only 19% chose to rely wholly on their own endeavours, less than the number (25%) that said they relied entirely on the library for their on-line searches (Table 5.21). The rest - the large majority - were quite content to do both. Just because large numbers of journalists could - and did - search for themselves, this did not mean that the library had no on-line role to play. This is heartening news for those librarians who had feared that end-use would sound the death knell of intermediary searching, though it has to be said that, the fact that journalists had restricted password access, must have played a part in boosting the delegation figures. We were also dealing here with only one on-line system: all NEXIS searching for instance was done by library staff, it was alleged by library staff that even PROFILE-proficient journalists felt that was beyond them - though not

apparently for Harman's (1986) Reuters journalists. There were, however, big differences between departments (Table 5.21). City journalists, with (significantly) only one librarian on call, generally preferred to search for themselves (56% of them said so), while in Features a large proportion (39%) of journalists entrusted all on-line searching to the Library. Generally, it was a case of the heavier and more proficient the searcher the more they searched themselves.

TABLE 5.21
EXTENT OF ON-LINE DELEGATION OF FT PROFILE
SEARCHES AT *THE GUARDIAN* (Survey)

		Wholly end users	Sometimes end users	Wholly delegators
City	n	5	2	2
	%	56	22	22
Features	n	0	14	9
	%	0	61	39
Foreign	n	3	4	1
	%	38	50	13
Home	n	4	10	6
	%	20	50	30
Sport	n	0	2	0
	%	0	100	0
Others	n	2	9	0
	%	18	82	0
Total	n	14	41	18
	%	19	56	25

NB. Users and non-users answered this question.

There were four main reasons why journalists delegated the on-line search. What encouraged most to delegate was the proficiency of the Library staff - this accounted for 31% of the reasons given (Table 5.22). Perhaps to that figure should be added those respondents, who, while not directly acknowledging the greater skill of the library staff - something that maybe journalists were reluctant to do, did say that they went to the library for difficult searches: this was the reason for delegation in 27% of the times. This might initially appear surprising, given the apparent simplicity of the system, but it has to be remembered that only a few journalists had been trained to use PROFILE and that the system is only simple in regard to its range of commands - getting something out of it can be difficult. The third and fourth reasons, respectively, were the problems encountered in getting a line out (26%) and lack of time (17%) - unexpectedly last, perhaps, given the enormous time pressures journalists were under (possibly, the act of delegation consumes more time, and takes longer, than doing it poorly yourself).

TABLE 5.22
REASONS FOR DELEGATING THE ON-LINE SEARCH
(Survey)

		Access problems	Lack of time	Search difficult	Librarian better	Total
City	n	0	0	1	1	4
	%	0	0	25	25	100
Features	n	10	5	8	11	36
	%	28	14	21	31	100
Foreign	n	2	1	1	3	7
	%	29	14	14	43	100
Home	n	7	5	9	10	32
	%	22	16	28	31	100
Sport	n	0	1	2	0	3
	%	0	33	67	0	100
Others	n	4	3	3	3	13
	%	31	23	23	23	100
Total	n	23	15	24	28	90
	%	26	17	27	31	100

NB. Some journalists chose more than one reason. Numbers exceed the population of respondents.

Journalists in the UK have traditionally conducted their own research. However, with on-line growing as a research tool and with a body of accomplished and trained on-line users in the shape of the librarians, an obvious (and related) question to pose was how willing were journalists to delegate some of their research to the Library to conduct on-line (Table 5.23A). A large majority (79%) were in fact willing to do this. And it was not simply the non-searching journalists who were willing, for even 46% of the heavy users indicated their willingness (Table 5.23B). However, it was true to say that the heaviest users were the least willing: 54% of them said so. Of *The Guardian's* departments, City - heavy users - were by far the least willing to delegate the research function: half their journalists were unwilling (Table 5.23A). The main reasons journalists gave for doing the research themselves (and at the same time providing further explanation as to why they cannot delegate the search) were:

(1) An inability to explain their information needs to a third party or, indeed, to themselves; it was only through browsing that they saw what they needed: typical responses were - *its usually difficult to specify precisely what l want, it's only by browsing do l spot what l need; I am looking for something and in the process often find something of similar or greater use; I have the most extraordinary tastes, and I would not expect anyone* [the librarian] *to second guess them; and I often find inspiration in information which I might not have initially though relevant.* Now, PROFILE caters well for this type of searching, for its' command driven interface provides an unencumbered workspace upon which the journalist can draw: no constraints or prods in the shape of a list of

prescribed choices, just **get** and **pick** and the unusual, but sometimes, interesting associations they create.

(2)   Related to (1), they wanted full control over the conduct and direction of the search. Delegation, it was felt, would cramp the creative process too much: this was the most frequently voiced reason for journalists conducting their own searches - *I prefer to follow my own line of thought, you are unable to make further checks on what is uncovered* [if you delegate]. And even elsewhere this seemed to be the principal reason for end use. Take the case of a *Times* journalist detailed to write a commemorative story on Bernard Levin (Erbach, 1994). On asking a librarian for information on how to access PROFILE, she suggested doing it on his behalf, but he declined the offer, saying that he did not know what angle he was going to take. By way of explanation he said that he would probably look up a couple of biographical portraits in other papers, zero in on something and then seek more information on that. Of course, the library had a bulging file on Levin, but this was too limiting in terms of its cross referencing abilities for the creative mind.

(3)   Insufficient time to delegate: characteristic of the responses here was this one - *takes too long to explain what l want*. Journalists are forever in a hurry.

(4)   The information required was too specialist or technical in nature - several City journalists were of this opinion. In the case of City journalists, they had just one librarian to delegate to - and that person was not a professionally trained librarian - you have to have confidence in the skills of the intermediary if you are to delegate. It was noticeable how journalists using the main library sought out specific staff to delegate their queries to.

Home news reporters were the most willing to entrust their on-line research to librarians (88% were willing): probably a reflection on the low-levels of on-line expertise in the department. Also, with having to cover widely varying subject areas, and the speed with which they have to work and the short time they have to collect information; they are probably only too happy to delegate. And more than any other department their journalists know and trust the library staff on an individual basis. Features journalists were also very willing to delegate - and as we have seen they were already extensive on-line delegators. This was surprising because feature writers are as proud of their research skills as they are of their writing skills, and often do not know what they are looking for until they see it.

TABLE 5.23A
GUARDIAN JOURNALISTS' WILLINGNESS TO DELEGATE
ON-LINE RESEARCH: ANALYSIS BY DEPARTMENT (Survey)

		Willing	Unwilling	Total
City	n	4	4	8
	%	50	50	100
Features	n	21	5	26
	%	81	19	100
Foreign	n	5	2	7
	%	71	29	100
Home	n	23	3	26
	%	88	12	100
Sport	n	2	1	3
	%	67	23	100
Others	n	8	2	10
	%	80	20	100
Total	n	63	17	80
	%	79	21	100

NB. Both on-line users and non-users answered this question

TABLE 5.23B
GUARDIAN JOURNALISTS' WILLINGNESS TO DELEGATE ON-LINE
RESEARCH: ANALYSIS BY TYPE OF ON-LINE SEARCHER
(Survey)

		Willing	Unwilling	Total
Non-user	n	24	1	25
	%	96	4	100
Occasional user	n	19	4	23
	%	83	17	100
Regular user	n	14	5	19
	%	74	26	100
Heavy user	n	6	7	13
	%	46	54	100
Total	n	63	17	80
	%	79	21	100

The two sets of findings on delegation - the delegation of the straightforward every day search and the delegation of research, basically reinforce and confirm each other, for while 19% of journalists only ever searched themselves, a similar proportion (21%) were unwilling to delegate on-line research. Also the actual number of people willing to delegate the search (63) was not far removed from the number that presently did so (59).

5.8 Changes over time

During a decade of monitoring on-line searching at *The Guardian* end-use has grown enormously. In 1986, before FT PROFILE was put on ATEX the figure end-use represented one-tenth of all on-line use. Shortly after PROFILE was placed on ATEX end-use reached one-third of the volume of intermediary searching and by the early 1990s parity had been reached. The very latest data - that for April 1995 - shows that end-users now account for two-thirds of FT PROFILE use. In tow have come other changes too. Journalists are now much more aware of what on-line is and what it can do. Indeed, frequently they couch there queries in on-line terms - asking librarians whether they can search PROFILE for this or that, or suggesting perhaps they could try NEXIS if they find nothing on PROFILE. There has also been a recent upsurge in the number of journalists asking the library to undertake research on their behalf: the library believes that this is because of greater awareness and an increasingly widely held belief on the part of journalists that librarians are better searchers - something the survey confirmed. What will prove most interesting though, is whether this trend continues, because since 1994, the library has had responsibility for *The Observer,* and is now servicing the needs of two sets of, sometimes, competing journalists. Journalists might fear that this compromises the confidentiality of their searches. For this reason journalists at News International were not actively encouraged to delegate the search - though they did so in large numbers (Erbach, 1994).

Attitudes have changed as well. When busy persuading journalists of the wisdom of searching themselves in the early days of on-line at *The Guardian*, an argument on the following lines would be frequently encountered: on-line would increase the journalists' burden - it would do nothing more than speed vast quantities of second-hand data to the already information-overloaded journalist. In turn, forcing the journalist into an unbearable existence of having to sit at the terminal all day creating stories from the computerised information feeds. No going out of the office, no gathering of the raw data themselves - and most journalists - even those that who largely compile lists - see themselves as being investigative and creative workers. They would argue further, that the public would not be interested in such second-hand news reporting. In some cases journalists put up this argument simply as a line of defence - hardened journalist hacks felt threatened by the changes that were occurring to their work environment, but in other cases there was a genuine fear that the quality of journalism itself was being threatened. However, you do not hear these arguments anymore, even the doubters are now building on-line into their daily information seeking routines.

In other aspects there has been very little change indeed, certainly when one considers that the end-user facility has passed its tenth anniversary. Thus essentially journalists still search just one system - FT PROFILE. This is in part due to the fact that PROFILE has met the expanding journalist appetite for on-line data by increasing source coverage and improving its currency performance over the years. It is also partly to do with journalists' unwillingness to learn and listen, and management's desire to keep a cap on end-user searching - after all the number of passwords has remained much the same despite doubling, tripling use levels. Systems have been introduced to journalists to journalists by the researcher - DIALOG and POLIS for instance, but largely to no avail. The occasional enthusiastic journalist has taken out a (usually short-lived) personal subscription to DIALOG or Compuserve, but that was about it. Currently, the Internet, to which most journalists have access, is making waves and this could presage a sea-change in the information seeking behaviour of journalists, though novelty and fashion could still be the driving forces here (the jury is still out on this). What is probably most surprising is that whole groups of journalists have remained untouched by on-line revolution - sports journalists and sub-editors have hardly ever searched PROFILE or any other on-line system for that matter.

There also has not been any real change in the searching skills on display. The knowledge of commands and their use, seems not to have altered much since the very beginnings of observation. The period between the survey (conducted in 1992) and transactional log analysis (1995) shows no changes in the level of searching skills. What seems to have happened is that more journalists have been drawn into the end-user net, but once in the net, they have shown very little inclination or need to rise up the ladder of expertise. And, of course, given that journalists are not subject to even the briefest of training programmes there is very little opportunity for them to improve their skills. The best searchers - the City journalists - have always been the best: the City journalists were as knowledgeable as some librarians back in 1986 - and this still remains very much the case. No amount of exposure to on-line - journalists are competitive and must have noticed that some of their colleagues search and obtain something useful for their endeavours - has changed the fundamental non-user stance of sub-editors of sports journalists for instance. Though, in the case of sports journalists anyway, things might be changing for they are beginning to take an interest in on-line searching - a new staff member recruited from the general news world and familiar with on-line is setting an example and the recent influx of tabloids on to PROFILE could be having an influence too. Perhaps the department is not wholly convinced for they sent their secretary to the library for training.

Other uses of on-line have also been trailed in front of journalists, but have failed to grab their interest. Selective dissemination of information was demonstrated to specialist journalists in 1987 - the medical, South African and social services correspondents took part in a brief experiment (Nicholas *et al* 1987: 97-108), but the nature of the job, where no one topic remains interesting for long (it wouldn't be news if it did); the fact that journalists want to trawl such a wide patch, and the fact that when the facility was floated to journalists (late Eighties) the existing SDI services could not produce sufficiently up-to-date information; meant that it came to nothing. As reported earlier though, the South African correspondent has taken out a SDI subscription with Compuserve, so things could be changing here too. The use of on-line as an investigative medium was also demonstrated: in this case the particular use of POLIS in tracking Members' interests/lobbies. Political and lobby journalists initially appeared interested, but either the fear of the new or a general displeasure with being shown by an outsider how to research their own field, meant nothing again materialised. It has to be said that some journalists, most notably ones from the City Desk have always used on-line creatively.

5.9 Conclusions

What then can be concluded from over ten years of monitoring *Guardian* journalists' on-line searches? Firstly, that journalists did conform in many ways to the stereotypical picture that we have of the end-user. Secondly, but not in every way - and there does require some revision in the way that we characterise them. Thirdly, that there were large differences amongst end-users in their on-line behaviour. The principal findings can be summarised as follows:

- If the need for rapid information retrieval is there - as it patently is with journalists - end-use can flourish even in the most trying of conditions (and there cannot be many more trying conditions than sharing two passwords - and work deadlines - with a hundred plus of your colleagues). A large number of journalists now regularly conduct their own on-line searches, though this probably still constitutes only a quarter of all *Guardian* journalists. The proportion has hardly risen over the last three years, so this could well represent the high tide of end-user searching at *The Guardian* - under current conditions at least. However, there appears no waning of interest - and for many journalists it is very much a case of once bitten ... . Indeed, journalists want more and better on-line access: this was very evident in the high demand for improved access to PROFILE and on-line training. What was not so clear is whether they could handle another host, though

it is possible that the most accomplished searchers like, City journalists, could. No newspaper that we know of offers more than one host for end-user searching. Of the cadre of on-line active journalists, for most, searching was a daily activity - something probably dictated by the newspaper's daily cycle. A significant number did achieve volume levels normally associated with information professionals, searching five, six or even seven times a day. On-line use was not spread evenly throughout *The Guardian's* several hundred journalists, City journalists were especially active and Sports journalists and sub-editors - despite sitting in close proximity - hardly searched at all. It appears that end-use appeals to those journalists that are: heavily dependent on current data, well catered for on-line and already independent information seekers. **Some journalists were very active on-line, but there were big differences between groups of journalists. In this there was no stereotype.**

- Journalists' searches were generally productive, with only 18% of their searches failing to find anything, and almost as productive as those of the librarians - 17% of their searches failed to find anything. Zero hit searches aside, there was a difference between the two groups, for the searches of journalists typically produced either a very few references (ten or less) or lots of them (more than forty), whereas the searches of librarians tended to fall between these two bands. The polarisation in the journalists' searches can probably be put down to the two very different types of search that they conduct: the fact-checking search, typically associated with an individual or company, and the broad subject - *keeping my options open* - sweep. In the case of the librarians, it was almost as if they thought that, for a search to be good, it must produce a standard number of documents - and that number should, at one and the same time, be manageable and yet still demonstrate their on-line prowess, especially in regard to recall. Nearly two-thirds of journalists' searches ended up in full-text displays - another possible sign of search success. High volumes of use also provide a testament to the success of searches, for journalists have so many resources available to them and with so little time to spare they certainly would not use on-line unless it produced the results. **This finding shows that it is too simplistic to think of end-users as unproductive and unsuccessful searchers and intermediaries as productive and successful searchers. Though librarians were generally more satisfied with their searches than the journalists.**

- Logically, it might have thought that the heavy on-line users would be more successful with their searches. But this was not really borne out by the data

because, although slightly more heavy users said that they were nearly always successful with their searches, fewer moderate users were dissatisfied with their searches. There were probably two explanations for this: firstly, high levels of use do not necessarily guarantee success in searching, although they seem to increase proficiency; secondly, the high volume, proficient searchers were that more critical of their searching. **Perceptions of success are not necessarily enhanced by high volume searching.**

- Journalists certainly browsed through a lot of material on-line: nearly one-third of searches displayed more than 50 documents and 43% of their searches saw the display of more than 10 screens of data, but this should be put into perspective, for it was a full-text system that was being searched and librarians viewed even more screens per search - much more, if we allow for the under reporting of their data because of their frequent use of the **nobreak** command. The idea, that journalists, forced by poor searching skills to scroll through screen after screen to unearth something relevant is then a nonsense. Indeed, it is the librarians that did this, partly because of their professional tenacity, partly because of an obligation to give users more data, so that they could fine tune it, partly because they were not so well placed to make immediate value judgements and partly to do with the fact that the librarians get a lot of the more difficult searches delegated to them. It is sometimes said that journalists, especially features journalists, like wallowing in information, much like a rhinoceros likes wallowing in water (Kelly, 1995). However, this study provides no evidence to suggest that this was done in connection with on-line. On the basis of full-text records displayed, number of records retrieved and screens displayed it was often the librarians that did the wallowing. **This is another finding that would challenge the stereotypical picture - there were no general signs that vast amounts of data being displayed or abandoned was the order of the day for end user searching.**

- Journalists' searches were extremely short - most conducted well within 10 minutes, 4 minutes being a typical time. In fact their searches were accomplished (a little) more quickly than those undertaken by information professionals. For a largely untrained user group, using a full-text database with a cumbersome **get** and **pick** approach, this must surely come as a surprise: intuition would surely have suggested a higher figure. The brevity of their searches can be put down to a variety of factors: journalists were better placed to make relevance judgements so could skim through data more quickly; they were always in a hurry (*journalists are notoriously impatient* - Erskine, 1994: np) delegation takes time so they

conduct the hurried searches themselves); there was limited interaction with the data; they used only a limited range of terms and commands. But what is fast or quick searching a sign of? It is too easy to fall in the trap and see speed as a component of skill. Plainly, this was not the case for speed of searching was allied to the (short) amounts of time available. The one thing that can be said for certain, is that quick searches are cheaper. **You cannot call end-users slow searchers - indeed sometimes they were faster than the information professionals. You can call them honest searchers though - the estimates of how much time they spent searching given in the questionnaires tallied with the figures given in the computer logs.**

- Overall journalists' searches ranged over a wide number of sources, a wider range in fact than was searched by the librarians. Librarians, of course, had another host to search (NEXIS) and had their own in-house resources; journalists on the other hand invested all in PROFILE. The situation was also complicated by the fact that, in both groups' cases, there was a marked preference for just one file, UKNEWS, which is in fact a multi-file group, consisting of 28 newspapers including *The Guardian*. What was clear though, was that journalists were not using on-line to just access their own newspaper: something that proponents of in-house databases need to bear in mind. However, what journalists seldom did was to change file once on-line (they changed file just one in ten times, in fact); something that was probably not so important when you were searching UKNEWS. But in this journalist were no different to librarians - they changed file with the same frequency. **End-users were not notably single source oriented, no more so than librarians, anyway, but then they did only have PROFILE to search. This is another finding that challenges the stereotype.**

- Journalists frequently conducted big retrospective sweeps when on-line, confounding the image people have of journalists as being preoccupied with solely the current situation. To believe this is to misunderstand how journalists use on-line, for they use it typically to obtain background data for stories they are currently working on - stories they have typically been alerted to by wire, fax or phone. Also, it is time rather than obsolescence that concerns them, if databases can speed you years back in time - and it takes no more than seconds - then why not go back? In more than half the cases journalists searched back to the file's inception - and that could mean searching back as much as 10 or 12 years. Some journalists indicated that they would like to go back even further than that. No doubt, as a result of the drive for greater recall or the research-orientation of the

questions they were delegated, librarians searched more regularly back in time. **End-users did search retrospectively - especially when it did not involve them changing file.**

- Journalists searches were typically straightforward and simple, in that they contained few terms and commands. The hallmark search was a stripped down or bare essentials one, which gave very early contact with the data, and then maximum time on-line to browse and display records. Many searchers were undertaking name searches, looking for biographical information - and this type of search was generally less problematical than the subject one, where alternative and synonymous terms had to be considered. That is not the case if biographical detail is required of a well-known person. Some of the problems end-users normally encounter when entering compound statements were circumvented by entering search statements in the form of a phrase - something that PROFILE made easy. But it was not simply a case of using phrases to bypass Boolean searching, for it did appear quite *natural* for journalists to express their queries in terms of phrases. They did not disassemble their queries into keywords, and then shepherd these terms, into facets and keywords, as librarians frequently did (the absence of the Boolean **or** and nesting suggests this), although the librarians, themselves, were also very partial to the phrase entry, in fact out scoring the journalists in this. The general impression obtained was a very similar one to that obtained by librarians at News International - the other major source of end users in the media - journalists were not so much bad searchers, as simple searchers (Erbach, 1994). **End users search statements were simple in design and implementation. This largely conforms to expectation.**

- Not surprisingly journalists proved on the whole to be subject searchers. The survey indicated that three quarters of their searches were of this type, with the logs indicating an even higher figure (85%). **Journalists chase information, rather than documents and this very much showed in their searches.**

- There was a hesitancy about some of the journalists' searches, and they were not so sure footed in the execution of the commands - the relatively large number of input errors provides evidence of this, though it has to be said that, librarians made even more errors, due in part it is believed, to the greater complexity of their searches and problems with backspacing. Although journalists' searches generally involved more steps than those of the librarians, this was largely because of their preference for the **pick command**, which by definition boosted the stage

count. Their searches only occasionally developed, in the sense that search strategies were adjusted in the light of what was found. While journalists' initial choice of terms showed a soundness and directness, the use of alternative words or word forms were very much on the low side. There were a few exceptions to the rule - mainly City desk journalists, but an end-user stereotypical picture did emerge. **End users did not appear to change their search strategies on the basis of what was found.**

- Journalists' use and knowledge of commands was poor: truncation, Boolean **or**, field and word proximity searches were barely practised - and it cannot be argued that any of these search facilities were peripheral to the use of FT PROFILE. Whilst journalists got away with using well below half of PROFILE's limited command range, librarians employed virtually the whole range. While a lack of training - and an inability to make the time for training - must have a lot to do with it, the fact that they were not always sure of what they were looking for, would inevitably result in simple and imprecise searching, which in turn required only a limited command range. The lack of command knowledge had the effect of both reducing recall (through the lack of truncation and the use of the Boolean **or**) and increasing it (through the lack of field searching). **End-users demonstrated a very limited range and grasp of search commands but often what they were looking for was complex in nature or difficult to undertake.**

- High levels of on-line use are no guarantee of search excellence and, equally, long years of searching does not necessarily provide the expected dividends, in the form of improved searching skills: observation, interview and logs show that there has been no general improvement in searching skills - they still remain very much at the elementary/novice stage in terms of command utilisation and query formulation. Searching a lot does not necessarily result in people developing their skills, extending their command range etc. Where journalists have expertise - and require little training, is in relevance assessment, and browsing through records. Journalists, accustomed to working through piles of documents in pursuit of a *newsworthy* story, think nothing of adopting this style of searching on-line: intuition and a good eye for a story are the major retrieval devices. Librarians too mimic journalists information seeking behaviour in order to give journalists what they want. There was a general recognition on the part of journalists that they did not have all the necessary on-line skills - the constant clamour for training seemed to indicate this - so they moved quickly to the on-line activity to which they could bring their skills to bare. This skill compensates for the lack of command

knowledge and helps to square the high levels of use and satisfaction with low levels of on-line expertise? **End-users were generally satisfied users, thanks largely to the fact they were better able at sifting through data in pursuit of relevance.**

- When searching styles and strategies of journalists and librarians were compared, what stood out most was that librarians looked more to increase recall, through the use of truncation, the Boolean or and synonyms and related terms. Of course, on a full-text natural language database, this type of searching brings with it an increased risk of noise - and perhaps one of the reasons why the librarians actually retrieved more records per search than journalists. This though was balanced by a greater use of field and proximity searching, which would have had the effect of increasing precision. It was possible that the reason why librarians produced more records per search was because they obtained more of the long and difficult searches - the research type ones. Journalists, on the other hand, with their *enter the first phrase I think of default approach* undertook (not knowingly) low recall searches, but because of their low use of field and proximity searching facilities, recall was not in fact as low as it might otherwise have been. Librarians were also more likely to persist in finding a better return for their searching effort: the logs showed that they did this by changing their search terms more often, or by returning to a search again, maybe 10 or 20 minutes later, after further consultations with the user. **In construction end users' searches were essentially low recall searches, but in execution they become broader in scope.**

- Despite searching a lot themselves, journalists still delegated much of the searching to the librarians, especially in connection with the more difficult searches. Delegation was undertaken least in the case of the most knowledgeable, high volume users - and City journalists were a case in point, but what has to be taken in account as well was that City journalists had less in the way of intermediary support, so they might have had little choice. Many journalists also indicated that they were willing to delegate to librarians some of their research activities. **End use had little impact on library searching, and if anything, helped to drive it even higher.**

- Most journalists were not trained in the use of PROFILE, but they did voice a need for training. But despite the apparently high demand for on-line training, when it actually came down to it, initial enthusiasm for training waned - and few journalists showed up for the training sessions. Their interest in on-line seems to

be pragmatic and there appears to be no obvious willingness or enthusiasm to get to grips with the finer points of searching, or indeed, with the source material available. A management-lead push for training might prove the answer, but this would be unlikely to happen, for they are generally too preoccupied with other things. **End users were not trained in the use of FT Profile, and liked the idea of training, but not the reality, of it.**

- The results of long exposure to on-line systems has resulted in increased volumes of end-user searching. The most recent evidence would suggest that on-line might have reached all the journalists that it is likely to reach under the present difficult conditions, for the proportion of journalists searching PROFILE has not increased over the past three years - so it must be a case of those journalists that do search, searching more. There was little evidence to suggest that journalists were using a greater command range or were becoming technically more proficient. **Exposure to or familiarity with on-line searching does not appear to result in the greater use of system commands.**

CHAPTER 6
CONCLUSION

Essentially what the study set out to determine was whether end-users outside of academe fitted the stereotypical picture painted of them. In particular, did they merit the largely uncomplimentary descriptions of their searching found in professional information circles? The study attempted to put the record straight through a detailed and wide-ranging evaluation of two key social science end-user groups - journalists and politicians. Librarians servicing these two groups were also studied to provide the essential contrast and reference points. A priority throughout the research was also to explain end-users' searching in the light of their information needs and wider information seeking behaviour. A subsidiary aim was to discover whether there were any similarities between the well documented searching practices of academic end-users and those of the not so well documented non-academic end-users (practitioners).

It is believed that this study of practitioners as end-users is one of the biggest of its kind conducted in the social sciences. Through these pages the searching behaviour of 170 or so end-users has been described and evaluated in the light of the searching behaviour of approximately 70 librarians. However, a degree of caution must be exercised when making generalisations or comparisons on the basis of the case study data. This is because the two studies featured different information systems, were conducted at different times and the methods and logs used did not always equate. For this very reason and because of the sheer sizes of the individual cases, it was thought best to conclude as we have gone along and substantial conclusions have been appended to each of the two case studies. Here instead we will: address the objectives set out at the beginning of the study; attempt to draw conclusions from both studies - noting where they reinforce themselves or disagree, but always bearing in mind their different origins; highlight data from either study that is thought to be especially significant; and relate our findings to those of other researchers.

High volume, satisfied searchers

Common to the two studies was the strong belief among both sets of librarians that users did not know what they were doing when they went on-line. To put it in its crudest terms end-users were thought to be ill-informed, clueless searchers. Neither group of librarians had any hard data to prove this though they were vociferous on this topic, recounting story after story illustrating the poverty of end-user searching -

and many of these anecdotal tales were tested in the text. Indeed, it was these very tales that had sparked an interest in doing the research in the first place, for there was always the suspicion in the mind of the researcher that some librarians were never that comfortable with the idea of end-use and never lost an opportunity to criticise it. What seemed to support the latter view was that, on the whole, end-users appeared satisfied with their searching. They demonstrated their satisfaction in three ways. Firstly, by searching the systems they were given access to - PROFILE, POLIS and TEXTLINE - heavily and in large numbers. The figures that really stand out here are: journalists accounting for nearly two-thirds of PROFILE use at *The Guardian*, with some of them searching the system more than five, six or seven times a day; and politicians' use of TEXTLINE accounting for 55% of all commercial database searching at The House of Commons, where the on-line searching Library staff exceeded 50 in number. And both groups had a choice in these matters, for they could always delegate to the professional librarians, who were highly trained, responsive and generally had excellent lines of communication with their users (in fact, in many respects, they were some of the modern professions best). Use was though not evenly distributed within the two organisations. Leaving aside MPs who seem to be something of a special case and considering their political staff instead, the on-line cadre that Richardson (1981) once spoke of amounts to about 20% of the users in both institutions. End-use is by no means universal, then.

Secondly, when asked in questionnaire or interview end-users said they were generally satisfied with their searches. In the case of *The Guardian* well over a third of journalists said they were almost always wholly satisfied with their searching and 53% sometimes satisfied (high praise indeed in the case of a system that was not easy to log into, and, when connected, not always easy to get relevant data out of: the librarians were more satisfied though with their searches) and at The House of Commons, where the question was phrased a little differently, politicians seemed even more satisfied - 84% were usually satisfied with their POLIS searches and 72% with their TEXTLINE searches. The lower level of satisfaction with the full-text system was probably due to the hit and miss nature of searching natural language, full text databases. Thirdly, end-user satisfaction was demonstrated by displaying lots of records from the respective systems. Research Assistants displayed nearly 100, 000 bibliographic records in the year surveyed and journalists displayed nearly 3,000 records - many full-text - in just two days.

How then do you reconcile high use and satisfaction with the librarians' observations? Well, in the first place, not every end-user was satisfied: 19% of respondents at *The Guardian* said they were rarely satisfied, and this rose to well over 40% in the case of female journalists. Librarians may have had these people in mind when recounting their stories, rather like a lecturer brands a class by the bad students they have in it, no matter how few in number. And certainly the Librarians at The House of Commons had their American researchers and *The Guardian* had its Home news reporters. Also, Lancaster *et al.* (1994) had a point when they said that, given the novelty and excitement brought to searching by on-line systems, end-users would say that wouldn't they. But they take it too far too discount all data on user-perceived satisfaction, preferring to seek refuge instead in systems-oriented, precision/recall scores. After all, for end-users on-line systems are not the be all and end all that they are to retrieval experts. End-users have many alternative lines of enquiry and what they say about on-line systems is set in this context: they are great at speeding back through vast quantities of data in pursuit of some query or inspiration. When it comes to system use its simply a question of horses for courses. It has also to be said that Lancaster *et al.*(ibid.) are probably confusing the expected outcomes of Librarians and end-users: why make comprehensiveness and recall the benchmark, when there are very good grounds for believing that practitioners at any rate, with their ever present concerns of currency, time and speed of delivery, prefer precision above all else.

Simple searchers

The two case studies provided ample evidence to support the widely held view (Peters, 1989) that end-users search simply, employing a very limited and basic range of commands in their searches. In the case of journalists they utilised barely 40% of the available PROFILE commands - ignoring even bread-and-butter commands like truncation and the Boolean **or**. By contrast *Guardian* librarians availed themselves of more than 90% of PROFILE's commands. They were also not just blinkered in the use of system commands but also in the fields they searched/specified. Thus of the fields specified in politicians' searches just three accounted for 89% of mentions, whereas in the case of librarians it took more than a dozen fields to reach the same percentage. And blinkered too, in the terms they used to form their searches - two-thirds of journalists' search statements consisted of no more than two words, for the librarians the equivalent figure was just over a half. Much of this though can be ascribed to lack of training. Journalists picked up their knowledge from their colleagues and on a trial and error basis when on-line. In these circumstances it could have been expected that just the essential key commands would be absorbed and that searches would be simple in construction. At the House of Commons too, end-user

command usage was much more limited than that of the librarians, but then the menus restricted their choice anyway. There was also a need to know element in their on-line searching: only so much time could be spared given the busyness and fragmentation of the working day. The preferred method of information seeking must have also had an influence on the low take-up. On numerous occasions individual journalists and politicians reminded the researcher that, because they did not always know what they were looking for until they stumbled across it, browsing was the much preferred on-line strategy. By definition almost, such searching is going to be light in construction.

Bad searchers?

Searching simply of course does not necessarily mean searching badly, although the association is there in much end-user writing. Similarly, utilising all the system's commands does not necessarily lead to success, and nor does the provision of sophisticated 'behind the scenes' search facilities like thesaurus intervention, synonym switching, or query formulation support: indeed, in the latter's case, it has been found that this can lead to users being side-tracked and conducting less successful searches (Robertson, 1995). If your information needs are hazy - as journalists and politicians frequently are, then simple searching is what one might expect. With search commands being largely abandoned the complexities of the search process are left to the document display stage, for that is were the search is refined, but that is largely done in the mind and not recorded on the logs. Given a choice of whether to search themselves or delegate the search - and most of the users featured here did both, it seems quite logical that it would be the more complicated and easily defined (this is not a contradiction in terms) searches that would be delegated (both librarians and end-users told us that). If this is true then we might expect librarians to employ more commands and fields in their searches.

In regard to the two indicators commonly used to measure search success - zero hit searches and errors made - end-users did perform less well than the librarians. In regard to the former: 24% of the searches of MPs' Research Assistants ended up with no displays of records, whereas the figure for HOC librarians was 18%; in the case of *The Guardian* it was much closer than that, respectively, 18% (journalists) and 17% (librarians). However, in the politician's study, where individual Library searching groups could be identified, there were a number of groups who recorded an even greater zero display score. Only a qualified pat on the back for librarians in this regard then. As to errors made in the input or construction of a search - and this data was only available for *The Guardian*, librarians made proportionally fewer mistakes in their use of system commands - 9% of their searches contained such errors, to the

end-users 20%; but then they made more spelling and input errors - 65% of searches contained errors to the end-users 44% (high levels of error occasioned by hurried searching).

Subject searchers

For politicians and journalists subject searching was very much the order of the day. This you would expect from systems like PROFILE and TEXTLINE, whose prime task, like that of politicians and journalists, was not document location - after all authors and titles and class marks are often nothing more than signposts to documents. In the case of PROFILE there was a smattering of date searches - date being especially significant to journalists, and the very occasional by-line search - but, having been regaled with tales of journalists keeping tabs on their competitors, far fewer than anticipated. Essentially though, well over 90% of searches were on subject, amongst these searches 'name as a subject' searches featured strongly. Phrase searching was very evident too. Even in the case of the bibliographic system, POLIS, subject searching predominated - 78% of searches were of this type. Rather in the same way as more by-line searches were expected of journalists, so more searches on MPs' names were expected of the politicians. But in fact they were more notable by their absence, occurring only 2% of the times. Conventional author searching at The Commons did not even figure in the top ten fields searched. Apart from the fact that politicians simply do not look for material that way, the other possible reason for rejecting the author approach is that the method of citation is complicated (issues of inversion, punctuation and abbreviation are all thrown at the user) and that makes for fiddly searching, though in the case of end-users, menus did provide some guidance. At The Commons the menus had been set up with the document form searcher in mind, a searcher which the system designers had assumed were in the majority. How wrong they were - form was only specified 16% of the times a field was searched for, and despite the advertising effect of the menus.

Quick searchers

To the researcher, probably, the most surprising finding was how quick end-user searches were. In some cases - that of journalists, for instance, the searches were astonishingly quick, with 4 minutes being the most commonly recorded search time. Leaving aside the in-house end-users of The Commons, for their logs could not be wholly relied on, both journalists (73%) and politicians (64% - POLIS; 54% TEXTLINE) accomplished the majority of their searches in less than 10 minutes - no doubt, a statistic that Commercial Hosts were well aware of when they introduced line and record charges. The speed of the search can be best explained by the urgency with

which both groups required information and the limited time in which they had to gather it. The longer time spent on-line by TEXTLINE users can be best explained by the fact that, because of the restrictions placed on access to the Main library at Westminster (the main source of desperately needed current affairs data) and an absence of alternative sources, it loomed larger in the life of its users. The fact that both groups rarely wanted comprehensiveness helped here too. There was very little evidence to support the anecdotal tales of end-users hogging the terminals or searching all day long. Indeed, in this light, they came over as highly disciplined searchers, posing no real financial threat to management. In both cases librarians did turn out to be marginally the faster searchers and this can probably be attributed to their greater skill and the fact that they get more practise. The speed figures probably do not do full justice to the skills of the librarians because there was some evidence to suggest they were delegated the more difficult searches.

Big browsers and large displayers

On the whole the logs and the surveys provided some support for the view that end-users browsed through large quantities of data in pursuit of something of value, but what was also clear was that librarians did this also: certainly end-users were not more conspicuous in this than librarians. Librarians, too, must adopt similar search strategies to the end-user, for they are familiar with the characteristics of their information needs and would know what was required in terms of search technique to obtain the required data. Stories of desk-side waste bins overflowing with rejected print-out or screens scrolling around endlessly can be consigned to the realms of folklore - no doubt from where these stories came; for with search times of 4, 5 or 6 minutes on-line - in many cases to full-text systems - there was very little scope for extensive browsing or displaying. For instance an average POLIS search for dial-up users displayed 27 bibliographic records and even the browsing journalist saw an average 58 records, of which 6 were full-text displays. The averages do camouflage the fact that in PROFILE's case, for instance, 16% of the searches saw the display of more than a hundred records - and these were probably the instances that librarians were thinking of when they make their criticisms.

Multi-file, multi-host users

Contrary to expectation, both groups of end-users utilised a greater range of files on the systems that they searched than their respective librarians. Their source selection was not as blinkered as might have been thought. There is probably a very good explanation for this: the librarians had other on-line sources to which they could turn, whereas end-users had to milk the limited range of systems that they were given

access to for all that they were worth. Even so, end-users at The House of Commons, who had access to two systems - as compared to the journalists one - searched more files or spread their searching time more evenly amongst the available files in the case of both systems. It would have been interesting to see what the situation would have been if end-users had access to the same number of hosts as their librarians - in neither of the featured institutions were they given that facility. One of the principal reasons for not giving them more hosts to access, was because they would not be able to handle the differences in search language/interface. This argument lost some of its credibility in the case of the politicians, for they proved adept at searching both POLIS and TEXTLINE. *The Guardian* management has flirted with the idea of providing journalists with access to NEXIS, and on the basis of the politicians' experience, journalists too would probably take to another system with some alacrity.

Retrospective searchers

No matter the great current concerns of both groups of end-users, they were still interested in searching databases some years back in time. In the case of the journalists where this feature was most evident - one quarter of the journalists said they searched back more than two years, and half of all the searches that were monitored by the logs were conducted on the whole on-line archive - amounting to more than ten years in some instances. In fact, on-line is a source of archival data when compared to wire service or twenty-four hours cable television news service, and that is how many people use it. The lines are however blurring, with newspapers on the database before they get to the news vendor (FT PROFILE) and wire service being updated every half an hour (DIALOG).

Menu-driven or command driven searchers

It was plain from the very heavy use of PROFILE by journalists that command-driven systems do not necessarily hold any fears for end-users. User-friendliness, if at all a determinant of on-line take-up amongst users - and this study would indicate that its role is exaggerated, is of course relative: you can have command-driven systems like PROFILE that are easy to use and menu-driven systems like TEXTLINE that are certainly not. A quarter of Research Assistants said they had problems with the TEXTLINE menus. It was also found that in the case of POLIS that the main searches approaches offered by the menus failed to cater for the end-users most frequent search - the subject one. In these circumstances menu-driven systems create as many problems as they solve for end-users, and when given the choice - as was the case with POLIS dial-up users - they sometimes migrate to the command-driven version of the system.

Differences amongst end-users

Perhaps, one of the most significant findings to come out of the study was the big, sometimes enormous, differences in searching behaviour between the principal end-user groups. This was particularly the case with politicians where the method of accessing the system - dial-up and in-house - produced such different results. In virtually all of the analyses dial-up users were the odd ones out - and it is believed that this cannot always be put down to the greater authenticity of the dial-up logs. Dial-up users' searches were undertaken much more quickly; they experienced more failures at the terminal; they displayed the most records per connect hour; their searching was spread out more evenly through the day and week (though not the year); they were the most single-file oriented; they used the least number of fields in their searches; they used the controlled language approach much more frequently than their in-house colleagues; and, by implication, used the command driven POLIS more.

In the case of political end-users there were really not any special factors to explain this, although Library staff were of the opinion that the dial-up users were a more dedicated group of users, simply by virtue of making the arrangements for dial-up access in the first place. The evidence on this though is unconvincing. Instead one needs to look for an explanation in the disciplines that come with searching on-line systems via a telephone line and an all-purpose PC work station. Also, it was quite likely that politicians used the dial-up facilities differently: using them for quick queries where speed of delivery was all important, and using the in-house facilities for the longer research-type queries. Support for the first explanation comes from another group of users elsewhere in The House: the Public Information Office staff, a group of on-line searchers staffing a telephone based enquiry service to the public - and whose policy was to answer a question without having to ring the enquirer back, had the most in common with the dial-up end-users.

With journalists the subject covered appeared to be one of the key distinguishing characteristics. Financial (City-desk) journalists were, individual-by-individual, heavier and more proficient searchers than their *Guardian* colleagues. They had an urgent and regular need for up-to-date financial information - and FT PROFILE has much to offer in this respect. As a result they were the early on-line leaders, and have continued to lead the way. It is also generally easier to search for financial information - the terminology is less vague and ambiguous, so financial journalists are likely to be more satisfied with their searches - as they were. But its more than all that

because there is a culture in the department, too, in which published on-line information is rated very highly.

Not so much end-users as untrained users

Training issues have preoccupied both studies. It has already been mentioned that in neither case were end-users given anything remotely like what an information professional would regard as on-line training. Neither were users provided on an individual basis with the most rudimentary searching guides. So searching behaviour was as much a function of untutored use as end-use. In the circumstances could anything other than simple, error-ridden searching be expected? The fact that this was not always the case is the real surprise. Librarians seemed quite quick to denigrate searching behaviour but did not seem so quick to provide the most obvious solution to the problem - training. This lack of concern or support could not be simply fobbed of with the excuse that the systems were user-friendly and, as such, required no training or support. For, while anyone can search PROFILE badly, it is in fact a very difficult and frustrating system to search well. It is beguiling in its simplicity, but to get it to work efficiently for you requires the utilisation of a broad range of its commands. Similarly, the so-called user-friendly menus of POLIS and TEXTLINE were fraught with problems - and the researcher defies anybody to say that either is easy to search - they are much more complicated to search than the average Library OPACs.

One of the on-line determinants that was put under the microscope throughout the study was the level of training and professionalism of the on-line user. This was especially so in the study of politicians, where it was possible to identify relatively small and discrete groups of Library staff and compare these groups with each other and with end-users. Indeed, what this proved was that professionalism and training were key determinants in system use, for the searching of the Oriel Room staff, The House of Commons library's most highly trained group (all professional librarians, who had years of experience in the Indexing Unit) stood head and shoulders above the rest. Whatever use measure might be chosen - documents displayed, search sessions, connect time - they searched proportionally more frequently and heavily than any other comparable group. This might have been expected for, after all, the system was designed with their needs in mind - they played the key role in providing parliamentary information to MPs. But their general professionalism showed in other ways too: they employed the widest range of fields of all groups; they experienced the fewest zero-hit searches; the telephone-based users aside, their searches produced more documents per hour; and their searches were the busiest, containing the most transactions. Of course, it could be argued, that all this was simply a function of their

heavy use of the system as it appeared to be the case with City journalists. Proficiency and frequent use appear to go hand-in-hand.

What was also evident about The House of Commons study was that there was so much variability in the searching characteristics of Library staff that you could always find a Library searching group that had some search characteristic in common with an end-user group. Thus some Library groups used controlled language less frequently than dial-up end-users, took longer to conduct their searches and experienced more zero hits. In fact there was often greater differences between Library groups than between Library staff and end-users. This again points to the danger of making sweeping generalisations about broad categories of on-line users. We have already seen how end-users themselves differed in their searching behaviour, and now heard very much the same about librarians. The differences between end-users was largely put down to the method of access, but in the case of the Library staff it was more complex than that: with training levels, degrees of professionalism; subject brief; computer literacy; relevance of the on-line system to work routines, and, perhaps most interestingly of all, a hard-copy culture, all playing their part in accounting for differences.

The time when searches were conducted

At the House of Commons the computer logs furnished on-line data for a full year. This enabled the pattern of searching to be examined over a large time span. While there was not a lot of variation in searching during the working day, there was considerable volatility during the days of the week and months of the year. The ebbs and flows of searching were very much tied to work rhythms, building during the week to Tuesdays Question time - the high point of weekly activity in The House and building during the year to the busiest legislative period in the Spring. This does throw into question the wisdom of asking questions about how many times a week searching was conducted: for searching can obviously be heavy and irregular. And this is probably the norm for many groups, with the possible exception of journalists, who are tied to a daily or weekly work cycle. For *Guardian* searchers there was only time of day data, but this nevertheless proved very interesting, showing what appears to be a change in the information seeking behaviour of journalists: they are doing more searching now in the mornings to prepare themselves for the day ahead, rather than as in the past, doing their searching at the last minute after a long day in the field.

Information needs

The key features of end-user searching were very much a function of information needs. The information needs of both groups of practitioners are characterised by a high degree of currency, the rapid retrieval of information, a secretiveness and an inability to always articulate the need. Furthermore information seeking is considerably constrained by a lack of time, very short and abrupt deadlines and a fragmented working day. In terms of on-line searching behaviour this manifests itself in the following ways: high volumes of end-use, frequent searching (especially in the case of journalists), quick searches, a high failure rate, a lack of command knowledge, a basic and simple searching style, a preference for viewing documents rather than assembling search statements.

Constraints placed on end-use

Both groups of users had big obstacles put in their way and yet end-use still flourished. For both groups there were the twin high hurdles of poor access and training. In theory, all journalists had desk-top access to PROFILE but in practice - with just two passwords between them - they found it very difficult to get a line. Politicians, too, were handicapped with having to share POLIS and TEXTLINE terminals amongst several hundred of them. And if that was not bad enough, for many users these terminals were sited in another building, about half a mile distant. As for training, the number of journalists who were trained on PROFILE could be counted on one hand - and that limited training given eight years ago. For politicians on-line training was a five minute, whirlwind Library tour: training was not thought necessary given the menu-driven nature of the systems provided and the opportunities for delegation in the case of difficult searches. Plainly then, poor access and no training are not going to stop end-use, though undoubtedly they will limit its growth - and this must have been what lay behind the thinking of both sets of managers - the editors of *The Guardian* and The House of Commons Librarian.

Delegation and non-use

Although high levels of end-use were recorded in the cases of politicians and journalists that is not to say that everybody wished to search for themselves. At *The Guardian* for instance three quarters of journalists delegate all or some of their searching. The reasons for delegating the search to the library were very much the same in both institutions, with the expertise of the library staff being cited most frequently. The reasons why politicians and journalist chose to search themselves were tied closely to their information needs. The major reason was not being able to frame a search statement in the first place, because they were not really sure of what they

were looking for - and the act of delegation and searching at one remove required them to be specific. Determining relevance was something that could not be delegated in these circumstances. The fact that users were also looking for an angle or slanted information (in the case of politicians) was another reason for searching themselves - and both groups were very much active in the processing and packaging of data. Confidentiality concerns were also significant, which made politicians especially insistent upon searching themselves - well over three quarters said that it was necessary for them to search themselves. Both studies showed that high levels of delegation and end-use could co-exist, with little threat to each - and in many cases users practised both, though heavy users, almost by definition, delegated less.

Given the fact that MPs' Research Assistants acted as intermediaries themselves - and sometimes to more than one MP - it might have been expected that they would search on-line databases more frequently than journalists, who did not employ researchers. The differences in the logs meant that we could not scientifically test this assertion. However it is estimated that the average dial-up Assistant spent: 4.75 hours on-line to POLIS during the year and said they searched the system a couple or so times a week, and 6 hours on TEXTLINE and searched it about weekly; whereas the average on-line searching journalist probably spent 75 hours on-line during the year and searched the system almost daily. The fact that this was not the case was probably due to the fact that journalists: had better access to the resource; were information providers themselves; and, typically had daily work deadlines.

In both case studies there were groups who remained untouched by the spread of end-user searching within their organisations. This was most evident in the case of MPs, where it was not possible to concretely identify a single MP as being an end-user. The fact that MPs were so little touched by on-line proved, initially, to be something of a shock, but was easily explained in the light of their broader information seeking behaviour and the time constraints placed upon them. But they also lacked a computer culture and did little, if any, on-line searching themselves. Indeed, many were hard put to name the House of Commons Library's own database, POLIS. They were aware of the terminals Library staff and their Research Assistants gazed into and had noted the increasing amount of print-outs they received from the Library - and might even, like several Members spoken to, expressed dissatisfaction about this trend (amongst MPs there was a lawyer-like suspicion about print-outs - for instance, were they original, authentic, to be trusted?). That was basically the level of their interaction with on-line systems. With the growth of on-line services - and especially the Internet, and their gradual absorption in all levels of society and walks of life it

remains to be seen whether they can continue to absent themselves from on-line searching. In some ways things are changing - the introduction of the Parliamentary Data Video Network will ensure a closer proximity to on-line, but it still seems that it will be their Assistants who will be the principals involved - and they are avaricious database users, as this study has proved.

Amongst journalists end-use was not quite so sharply demarcated but even so, sub-editor and sports journalists hardly searched at all - and that in a building where they may be just metres from where their colleagues searched, perhaps, 4 or 5 times a day. The reasons for non-use were somewhat different at *The Guardian*. In the case of the sub-editors it was not a question of databases being of no value to them - PROFILE is generally quite good for fact checking and alternative sources were in limited supply, what with the library being unstaffed when they typically did most of their cross-checking. It was a case of not thinking to use on-line: there was no realisation of its worth and that they had not been trained to value it. Perhaps in mitigation they do work to sharper deadlines than most other journalists - and this might well be another factor. As for sports journalists, their narrower field of view (e.g. Rugby), the fact that they are often enthusiasts (remembering Arsenal's every football score) and their informal information culture provides much of the explanation. Interestingly, the only sports journalist, who did search was an import from current affairs.

Changes over time

At *The Guardian* it was possible to investigate end-use over a long period of time - ten years in fact. This provided the opportunity to study change. Change was most obviously seen in the increase in end-use, spiralling ever upwards and now easily eclipsing that of the librarians: the on-line tables have undoubtedly turned. Changes were also detected in the more positive attitude shown towards on-line and the greater awareness of its potential: both contributory factors to the greater use of on-line. On the other hand, there was little evidence to suggest that searching skills had improved at all; that journalists were looking to do anything different with on-line (research, for instance), nor was there any strong demand for the introduction of new on-line hosts - though the Internet was sparking off great interest.

Complimentary methodological data

On a methodological note, computer logs and surveys provided complementary data and were, in fact, in harmony about many aspects of on-line searching. Proving that, in the case of practitioners anyway, surveys are a valid research instrument for the examination of end-users. The fear that respondents would butter up to the

information researcher or attempt to show themselves in the best light proved groundless. Journalists and politicians are very self-confident and do not need to curry favour with anyone. They plainly tell it as it is. This together with the fact that computer logs are increasingly being made available to researchers in the field means that the on-line world outside academia should be ripe for many more investigations of this kind. Obviously conducting research in a busy work environment cannot be compared to conducting it in laboratory or controlled conditions, and as a consequence the data is inevitably going to be more 'dirty'. How dirty is difficult to tell, but perhaps not that dirty, because , as we shall see, the results of this study are in the main in keeping with the results of studies conducted in much more controlled environments.

Still on the methodological front, at The House of Commons there was a reasonably high degree of agreement amongst the four use indicators: sessions conducted, connect time, records displayed and transactions made. They all told very much the same story: all pointing to the fact that one group of researchers were the busiest - The Oriel Room staff.

Comparison with previously published research

Finally, how do the findings described here tally with those of other researchers who have studied end-users? Reviewing the studies of academics first, we in fact find that there is generally a lot of common ground. We have already heard how there is a high level of consensus over end-users limited command use. The City journalists at *The Guardian* and The Oriel Room staff of The House of Commons Library were proof of Penniman's (1982) supposition that frequent users employed a greater command range. There was agreement too, over how economic search expressions were. Hancock-Beaulieu (1990A) discovered that her academics' searches averaged two terms - journalists searches were also characterised by having an average two terms. Levels of search satisfaction were very much of the order found by Peters (1989). There was, however, only limited support for Peters (1989) finding that a high proportion of searches apparently failed. 40% of the searches of academics monitored by Peters produced no hits, in the case of journalists and politicians they only failed to display documents on 18% and 24% of the times, respectively. There was also very strong support for Sloan (1986), who found big differences in on-line searching behaviour between in-house and dial-up users: students in their case; also there was corroboration for another Peters' (ibid.) finding, that a high proportion of searches were characterised by typographical and spelling errors - 44% of journalist searches were faulty on these grounds; and for Kirby and Miller (1986) who found the Boolean

or almost wholly absent from searches of end-users. The absence of truncation from students searches noted by Akeroyd (1990) is mirrored here in the case of both journalists and politicians. It follows that there is no corroboration for Tenopir *et al*'s (1990) observation that full-text users resorted to truncation quite frequently. Lancaster *et al* (1994) found that librarians conducted their searches twice as quickly as end-users, but we found that librarians were only a little quicker, although, to be fair, in the case of Lancaster *et al.* the comparison was made between users conducting the same searches. The following words of Murphy *et al* (1991) could easily have been said by this researcher when referring to journalists: *their level of expertise does not increase and they may never proceed to use more sophisticated facilities* [commands].

Where differences can be found in searching behaviour they can be attributed to the differences in information needs of the two groups. The information needs of practitioners are principally characterised by the need for currency and speed of delivery. This requires them to search frequently to keep in touch with rapidly unfolding events and to use rapid on-line retrieval methods whenever possible. The enormous time constants on their information seeking results in very short searches and the inability to articulate their need results in apparently simple subject searching. Also with academics you are dealing with a group for whom published documents have a special place and especially so students, who are given readings to track down. Inevitably much of their searching will involve bibliographic detail - author, title and class mark. But with practitioners, like journalists, hardly any of this obtains, so you would expect - and get - higher levels of subject searching. And even in the case of politicians, where they sometimes need to track down a White Paper or Hansard debate, the subject approach is ever present.

Of the practitioner studies, the early ones look distinctly dated. Citroen's (1984) view that end-users required extensive training before they could function now looks way out of line. There was no sign of the settling down stage envisaged by Richardson (1981), in both studies end-use has continued to grow inexorably, with growth having to be artificially held-down. Walton (1985) cannot have been thinking of politicians or journalists when he doubted whether end-users would ever search regularly enough to maintain their skills. Certainly there was no sign whatsoever of the demise of the professional on-line searcher prophesied by Jones (1984). Indeed, end-use appears to have increased the workload on the librarians of both organisations studied. Even with the more recent studies there is not a lot of agreement. Thus unlike Rigglesford's financiers, journalists did not search noticeably more broadly than librarians - they

used fewer terms, but that is not the same thing. More common ground can be found with Sullivan (1992) whose practising scientists' searches contained three or more steps - this was true for journalists too. Koblas' (1993) view that accessibility and ease of use are not prime determinants in end-use was very much borne out by the experiences of the politicians at The House of Commons, who had so many impediments placed in their way.

Closer to home in the media field, this study largely concurs with what has been said in the published literature on journalists. Journalists proved to be as poorly trained and supported as they were when Harman (1986) was investigating the on-line set up at Reuters, and Miles (1993) the situation at *The Times*, and yet they seemed no more willing to get trained than when Arundale and Erbach (1988) noted this. Allowing for the slightly different wording of the questions, satisfaction levels among *Guardian* journalists were very similar to that found amongst US journalists studied by Jacobson and Ullman (1989). Searching styles were very much the ones described by Miles (1993) at *The Times*. There were some important exceptions though. The random keyword search approach described by Stanbridge (1992) was not in evidence though, and nor did the absence of a facsimile of the news story prove an obstacle to searching (Poynder, 1993).

Finally, the limited number of studies that have been conducted on politicians were very much in line with the current research, disagreeing only significantly on the importance of full-text databases, like PROFILE, to MPs and their Research Assistants (House of Commons Information Committee, 1994).

Summary

To sum up then, in many ways end-users deserve their stereotype: simple searchers with a limited grasp of the commands. Much of this can be attributed to: (1) the fact that they were a largely untrained and unsupported workforce; (2) the way that they look for information generally. But they do not fit the stereotype in every way. There were some surprises: their speed of searching, their use of a variety of files/hosts, the alacrity with which they took to on-line, the tenacity they displayed in continuing to search when barriers were placed in the way, the extent to which they searched back in time, their searches did contain errors but so did those of the librarians, and while they certainly displayed large numbers of documents, so too did the librarians. Practitioners and academics end-users do have a lot in common when they search databases, but there were key differences and these were largely faster search times, a stronger subject search orientation and less of a need for comprehensiveness.

6.1 RECOMMENDATIONS AND IMPLICATIONS

As the findings of this research challenges some of the assumptions that have long been held by the information profession there are a number of recommendations and implications which should be noted. There is something in the research for all groups; and this is not surprising for after all they share a common constituency - the information user. However certain findings are especially important to some groups, and the section is organised along these lines.

INFORMATION MANAGERS AND TRAINERS

- While many end-users do conform to the stereotypical view held of them a sizeable minority do not. Typically this minority are frequent or heavy users - and this makes up for the absence of training, which is the most marked characteristic of end-users. Indeed, in many cases, we should read untrained user for end-user. Training rather than user-friendly interfaces should be the priority for information managers.

- Being a simple but satisfied searcher is not a contradiction in terms and information professionals should not fall into the trap of associating simple searches with poor searches. Furthermore, in making such judgements information professionals should view the entire search process and not neglect the display of documents, for it is at this stage that endusers compensate for the shortfall in their understanding of the system's search commands.

- Information managers need to give serious thought to the areas where the stereotypical picture does not conform, for there are major training and cost implications here. Practitioner endusers spend very little time on-line when they search and are not the financial liability many have thought. They search a long way back when given the opportunity and would be seriously inconvenienced by limited archives, like those offered by CD-ROM. The fact that they search over a wide number of sources, showing a marked preference for multi-file groups also shows up the limitations of the current CD-ROM technology. Endusers reduce the pitfalls of untutored searching by searching simply on specific terms, typically phrases - the move towards broad subject help menus should be seen in this light. Given that there was a high degree of consensus between the two groups (politicians and journalists) studied, it is quite likely that these searching traits are shared by many other groups of social science practitioners.

INFORMATION RESEARCHERS

- Some of the measures used by academic researchers to pigeonhole on-line searchers and to establish good and bad searching practice are highly dubious, sometimes leading to ridiculous statements that bare little understanding of the complexity and heterogeneity of the searching process. Transactional log analysis researchers are particularly guilty here. Logs give rise to countless opportunities for type-casting searches and searchers, but none of them are without reproach. Thus the idea that: speed of searching indicates anything other than proficiency; the more records found the more successful the search; a zero posting always denotes a failed search; the more commands employed somehow the more superior the search: is simplistic, and shows poor understanding of the information seeking behaviour of endusers. It is unlikely that on their own logs tell us anything sensible about the searching behaviour of academics, a relatively homogeneous group, and they certainly have even less to say about that of practitioners, who have little time to search or train, conduct a wide variety of searches and have many other information sources to turn to.

INFORMATION SYSTEMS DESIGNERS AND ON-LINE HOSTS

- With increasing amounts of log data available and an ever-eager group of people waiting to analyse it, there is a danger that such indices will become established and accepted by information systems designers. After all they are so easy to collect and the sheer amount of data that can be brought to bare has a Siren-like attraction. The log data itself is not at fault, it is what is read into it. Log data should always be treated with caution and used in conjunction with other forms of data, especially survey data. Indeed, on-line hosts should make their transactional log data available to researchers so that this can come about - they themselves are unlikely to conduct surveys. It is also important that when major information systems are introduced they should be accompanied by facilities for user-evaluation. This evaluation should be longitudinal in nature in order to view change and to allow searching styles to establish themselves.

- Furthermore, hosts should take cognisance of the length of endusers' retrospective sweeps before they consign 'historical' files to CD-ROM storage. They should not be so quick to introduce new, supposedly user friendly, interfaces - and sow further confusion as a result amongst an already confused population, without taking a much stronger role in the users' education and training.

REFERENCES[1]

Akeroyd, J. Information seeking in on-line catalogues. *Journal of Documentation* 46(1) March 1990, 33-52.

Arundale, J. *and* Erbach, G. The Library of the Independent: a case study in the introduction of information technology in a newspaper library. In *On-line Information 88: 12th International On-line Information Meeting, London 6-8 December 1988.* Vol. 1. London: Learned Information, 1988. 575-580

Barbuto, D. *and* Cellavos, E. End-user searching: program review and future prospects. *RQ* 31(2) Winter 1991, 214-227.

Barber, A. *and* Riccalton, C. *The use of LS2000 Online public access catalogue at Newcastle University Library.* London: British Library Research and Development Department, 1988.

Barker, C. *Quoted in*: Pros and cons of end-users' direct access to electronic data. *Information World Review CD-ROM.* March 1993, [np.]

Bath University Library. *Investigation of the information requirements of the social sciences. Report 1: Information requirements of researchers in the social sciences; Volume 1. Text.* Bath University, 1971.

Batten, M. (Financial Times) *Personal communication.* 1994

Bennett, J. Storage design for information retrieval: Scarrott's conjecture and Zipf's Law In *International Computer Symposium, 1975.* Amsterdam: North Holland, 1975, 233-237.

Bodtke-Roberts, A. Faculty end-user searching of BIOSIS. In *National Online Meeting Proceedings.* New York: Learned Information, 1983. 45-56.

Bonta, B. Online searching in the reference room. *Library Trends.* 31, 1983, 495-470.

---

[1] *Note: some references sourced from full-text CD-ROMs, for which no page numbers available*

Borgman, C. *et al.* Children's use of an interactive catalog of science materials <u>In</u> *ASIS '90: Proceedings of the 53rd ASIS annual meeting.* Medford, NJ: Learned Information, 1991.

Bourne, C., Robinson, J., *and* Todd, J. *Analyses of ERIC on-line searching procedures and guidelines for searching.* Berkley, CA: University of California, 1974.

Brewster, P. Letter. *Information World Review CD-ROM.* May 1995.[np.]

Brittain, J.M. *The social sciences: the supply and demand for documentation and data: a report for UNESCO.* Rossendale, 1982.

Citroen, C. The end-user foots the bill - so what? *8th International Online Information Meeting Proceedings.* London: Learned Information, 1984.

Dalrymple, P. Clinical uses of MEDLINE on CD-ROM <u>In</u> *MEDLINE on CD-ROM* edited by R. Woodsman *et al.* Melford, N.J: Learned Information, 1989.

Dalrymple, P. *Retrieval by reformulation in two library catalogues: toward a cognitive model of searching behaviour.* PhD Thesis. Wisconsin; University of Wisconsin; 1987.

East, H. *and* Tilson, Y. *The liberated enduser: developments in practice and policy for database provision to the academic community.* London: British Library Board, 1993.

Ensor, P. *and* Curtis, R. Search Helper: low cost on-line searching in an academic library. *Reference Quarterly.* 23, 1984. 327-331.

Erbach, G. (News International) *Personal communication.* 1994

Erskine, R. *Quoted in*: Total recall for Mirror. *Information World Review CD-ROM.* May 1994, [np.]

Fenichel, C. *Online information retrieval: identification of measures that discriminate among users with different levels and types of experience.* PhD. Dissertation, Drexel University, 1979.

Fidel, R. Individual variability in online searching behaviour. *Proceedings of the American Society for Information Science.* 22, 1985, 69-72.

Fisher, J and Bjorner, S. Enabling online end-user searching: an expanding role for librarians. *Special Libraries* 85 Fall, 1994, 281-282.

Fjallbrant, N *et al.* End-user training in the use of a small Swedish database. *College and Research Libraries.* 44, 1983, 161-167.

Garman, N. End-users: the pot of gold at the end of the rainbow. *Online* 13(6) November 1989, 6-8.

Haines, J. Experiences in training end-users searchers. *Online.* 8, 1982, 15-29.

Hancock, M. Subject searching behaviour at the library catalogue and at the shelves: implications for on-line interactive catalogues. *Journal of Documentation.* 43, 1987, 303-321.

Hancock-Beaulieu, M. Evaluating the impact of an on-line catalogue on subject searching behaviour at the catalogue and at the shelves. *Journal of Documentation.* 46(4) December 1990(A), 318-338.

Hancock-Beaulieu, M. et al *Evaluation of on-line catalogues: an assessment of methods.* London; British Library Research and Development Department, 1990(B)

Harman, J. Reuters: a survey of end-user searching. *Aslib Proceedings* 38(1) January 1986, 35-42.

Harris, K., Nicholas, D. *and* Erbach, G. Information storage and retrieval in UK national newspapers: some effects of change. *Journal of Librarianship.* 19(2), April 1987, 71-88.

Harris, K., Nicholas, D. *and* Erbach, G. Online use and end-users in media and advertising: an overview. *Aslib Proceedings* 38 (11/12) November/December, 389-397.

Haygarth-Jackson, A. Online information handling: the user perspective. *Online Review* 7, 1983, 25-32.

Holmes, N. *Online information in Europe and the US.* Financial Times, 1985.

House of Commons. Information Committee. *The provision of a Parliamentary Data and Video Network.* London: HMSO, 1994.

House of Commons. Services Committee. *Computer services for Members.* London: HMSO, 1990.

House of Commons. Services Committee. *Information technology: Members requirements.* London: HMSO, 1984.

Jacobson, T. *and* Ullman, J. Commercial databases and reporting: opinions of newspaper journalists and librarians. *Newspaper Research Journal* 10(2), Winter 1989, 15-25.

Jones, K. The effects of expert and allied systems on information handling: some scenarios. *Aslib Proceedings.* 36, 1984, 213-217.

Kelly, S. *Information needs and information seeking behaviour of features journalists.* London: University of North London, 1995. (Unpublished Report)

Kirby, M. *and* Miller, N. MEDLINE searching on Colleague: reasons for failure or success of untrained end-users. *Medical Reference Services Quarterly.* 5(3), Fall 1986, 17-34.

Koblas, J. So why do people use on-line? An investigation into the discretionary use of electronic information resources. In *On-line Information 93: 17th International On-line Information Meeting Proceedings, London 7-9 December 1993.* London: Learned Information, 219-226.

Lancaster, F., Elzy, C., Zeter, M., Metzler, L., *and* Low, Y-W. Searching databases on CD-ROM: comparison of the results of end-user searching with results from two modes of searching by skilled intermediaries. *RQ* 33, Spring, 1994, 370-386

Larsen, G. End-user searching and the human aspect. In *On-line Information 88: 12th International On-line Information Meeting, London 6-8 December 1988.* Vol. 2. London: Learned Information, 1988. 467-474.

Larson, R. The decline of the subject searcher: long term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science.* 42(3), 1991, 197-215.

Leonard, T. Databases in the newsroom: computer-assisted reporting. *Online.* 16, May 1992, 62- 65.

Martell, V. The development of information technology for use by MPs at The House of Commons. MA Dissertation. University of North London, School of Information Studies. 1995-1996.

Martin, H. *and* Nicholas, D. End-users coming of age? Six years of end-user searching at **The Guardian**. *Online & CD-ROM Review,* 17(2), 83-90

Meadow, C. Letter. End-user searching. *Library Journal.* 108, 1622-1624.

Meadows, A. *and* Buckle, P. *A preliminary study of the scientific information system in the UK.* London: British Library, 1991. (BLRDD Report 6053)

Meister, D. *and* Sullivan, D. *Evaluation of users reactions to the prototype online information system, ERIC.* ERIC document ED019094. 1967.

Miles, E. *I.T. and the editorial production at The Times.* (MSc Dissertation) London: City University, 1993

Murphy, F *et al. Making OPAC interfaces to user needs.* Huddersfield: Huddersfield Polytechnic, 1991. (BLRDD Report 6041)

Nelson, M. Correlation of term usage and term indexing frequencies. *Information Processing and Management.* 24(5), 1988, 541-547.

Newkirk, J. *and* Jacobson, T. CD-ROM search strategy analysis: a pilot study. *Computers in Libraries Conference.* London: Meckler, 1993.

Nicholas, D. Letters: Systems driven at Wapping? *Library Association Record.* 95(4) April 1993, 222.

Nicholas, D. *and* Connolly, K. Big browsers are watching you ... . *Library Association Record.* 95(2) January 1993, 34-35.

Nicholas, D. *and* Connolly, K. Information technology developments in the newspaper industry and the future of the librarian. *Library Association Record.* 89(10) October 1987, 530-533.

Nicholas, D. *and* Connolly, K. To cut or not to cut ... and who does the cutting. *Library Association Record.* 95(2) February 1993, 104-105.

Nicholas, D, Erbach, G., Pang, Y. *and* Paalman, K. *End-users of online information systems: an analysis.* London: Mansell, 1988.

Nicholas, D. *and* Erbach, G. *Online information sources for business and current affairs.* Mansell, 1989. Chapter 1.

Nicholas, D., Harris,K. *and* Erbach, G. *Online searching: its impact on information users.* London: Mansell, 1987.

Nicholas, D., Harris, K., *and* Erbach, G. Time-Life, World Reporter and the secretary. *Journal of Information Science* 12, 1986, 167-175.

Nicholas, D. *and* Martin, H. Should journalist search themselves? (And what happens when they do?) in *On-line Information 93: 17th International On-line Information Meeting Proceedings, London 7-9 December 1993.* London: learned Information, 1993. 227-234

Nicholas, D. *and* Pandit, P. What happened to libraries in Independent Television? *New Library World.* 95(1114), 1994, 4-7.

Nicholas, D, Pang, Y. *and* Lievesley, G. *Online searching at The House of Commons.* London: British Library, 1991.

Oppenheim, C. Designing for the end-user marketplace. In *Information systems for end-users: research and development issues.* Edited By M. Hancock-Beaulieu. London: Taylor Graham, 1992 25-34.

Penniman, W. Modelling and evaluation of on-line user behaviour In *Information interaction: Proceedings of the 45th ASIS Annual Meeting.* White Plains, NY: Knowledge Industry Publications, 1982.

Perez, E. Newspaper libraries: automated and non-automated systems: electronic approaches. *Editor and Publisher,* 12 January 1980, 22.

Peters, T., Kurth, M., Flaherty, P., Sandore, B. *and* Kaske, N. Transaction log analysis (Special Issue). *Library High Tech* 11(2) 1993, 37-106.

Peters, T. When smart people fail: an analysis of the transaction log of an on-line public access catolog. *The Journal of Academic librarianship* 15(5) 1989, 267-273.

Poynder, R. Meeting the needs of journalists. *Information World Review.* June 1993, 8.

Quint, B. End-users vs. professional searchers. *Database Searcher* 4(8) September 1988, 4-5.

Re'em, D. Online searching. *Broadcast Technology International.* November, 1985, 40.

Richardson, R. End-user online searching in a high technology engineering environment. *Online,* 5, 44-57.

Rigglesford, D. CD-ROM: the answer for end-users? In *Information Systems for End-Users.* edited by M. Hancock-Beaulieu. London: Taylor Graham, 1992, 35-44.

Robertson, S. (City University) *Personal communication.* 1995.

Salovaara, I. Experiences in end-user and intermediary searching at the technical research centre of Finland (VTT) In *On-line Information 88: 12th International On-line Information Meeting, London 6 - 8 December 1988. Vol. 1.* London Learned Information, 1988. 103-110.

Saracevic, T., Mokros, H., Su, T. *and* Spink, A. Study of information seeking and retrieving. *Journal of the American society for information science.* 39, May 1988, 161-216.

Sloan, R. Hightech/Low profile: automation and the invisible patron. *Library Journal.* 111, November 1986, 4-6.

Standera, O. On-line retrieval systems: some observations on the user/system interface In *Proceedings of the 38th ASIS Annual Meeting.* Washington, DC: ASIS, 1975. 38-40.

Stanbridge, R. Journalists begin to embrace online databases themselves. *Information World Review.* 76, Dec. 1992, 46-48.

Sullivan, M. training for MEDLINE on CD-ROM: a case study in an industrial environment. In *Information Systems for End-Users.* edited by M. Hancock-Beaulieu. London: Taylor Graham, 1992, 71-78.

Sullivan, P. *Rhetoric and the search for externally stored knowledge: toward a computer-age art of research.* PhD Thesis. Carnegie-Mellon University, 1986

Tenopir, C., Nahl-Jakobovits, D., *and* Howard, D. Full text search strategies and modifications: the role of the searcher and the role of the system In *Beginning our second decade: proceedings of the Eleventh National Online Meeting.* Medford, NJ: Learned Information, 1990. 383-399.

Thompson, B. Future direct users of sci-tech electronic databases. *Database* 6, 1983, 6-9.

Tolle, J. Understanding patrons' use of online catalogs: transactional log analysis of the search method In *Productivity in the information age: proceedings of the 46th ASIS Annual Meeting.* White Plains, NY: Knowledge Industry Publications, 1983. 167-171.

Vreekamp, H. The information seeking attitudes of non-metropolitan journalists - a qualitative study of two communities and their primary and secondary sources. *The Electronic Journal* 13(1), February 1995, 43-51.

Walton, K. Experiences at Exxon in training end-users to search technical databases online. *Online* 7, 1983, 42-50.

Watson, M. Overdosing on information: the Parliamentary Data and Video Network. *Assignation* 11(2) January 1994, 27-29.

Zink, S. Monitoring users search success through transaction log analysis: the WolfPAC example. *Reference Services Review.* 19(1), Spring 1991, 49-56.

LIST OF PUBLICATIONS ARISING OUT OF
THE RESEARCH

Martin, H. *and* Nicholas, D. End-users coming of age? Six years of end-user searching at The Guardian. *Online & CD-ROM Review,* 17(2), 83-90

Nicholas, D. Are librarians really better searchers than end-users? (And whose story do you believe?) *Online Information 95: proceedings.* Learned Information, (Winter 1995)

Nicholas, D. The impact of information systems on user groups, with special reference to politicians and journalists. In *Information systems for end-users* edited by M Hancock-Beaulieu. Taylor Graham, 1992.

Nicholas, D. Letters: Systems driven at Wapping? *Library Association Record.* 95(4) April 1993, 222.

Nicholas, D *et al. Online searching at The House of Commons.* London: British Library, 1991.

Nicholas, D. *and* Connolly, K. Big browsers are watching you ... . *Library Association Record.* 95(2) January 1993, 34-35.

Nicholas, D. *and* Connolly, K. To cut or not to cut ... and who does the cutting. *Library Association Record.* 95(2) February 1993, 104-105.

Nicholas, D. *and* Martin, H. Should journalist search themselves? (And what happens when they do?) in *On-line Information 93: 17th International On-line Information Meeting Proceedings, London 7-9 December 1993.* London: Learned Information, 1993. 227-234

APPENDIX I

INTERVIEW SCHEDULE FOR MPs

Open-ended questions

Areas of questioning

1. Nature of job and typical working week (Diary entries).

2. General information needs: subject, nature, authority, viewpoint, currency, speed of delivery, nationality, processing/packaging.

3. Information seeking behaviour
Constraints on information seeking.
Role of the Research Assistant in all this

4. Sources of information

5. Use of information systems
Value and role of HOC library.

6. Major on use, non-use and general awareness of on-line systems and computerised information services. Attitudes towards.

7. Specific questioning on POLIS and TEXTLINE.

QUESTIONNAIRE FOR MPs' RESEARCH ASSISTANTS

SURVEY OF ON-LINE SEARCHING FACILITIES (POLIS AND TEXTLINE) AT THE HOUSE OF COMMONS

I   <u>Background Information</u>

Name: ....

MP(s) represented ....

Length of service with MP: ....

Age: [   ]   18-30   [   ]   31-40   [   ]   40 & over

Pass category: ....

II   <u>Use of Databases</u>
If you use only one of the on-line systems, please just tick the boxes in the appropriate column.

1.   How did you first learn about POLIS/Textline?

POLIS   Textline

[   ]   [   ]   from other researchers/secretaries

[   ]   [   ]   from the library staff

[   ]   [   ]   from your MP

[   ]   [   ]   other sources. Please specify

....

2.   Have you been trained in the use of POLIS/Textline?

POLIS   Textline

[   ]   [   ]   yes

[   ]   [   ]   no

[   ]   [   ]   no, but I would like to be trained.

3.   How often do you use POLIS/Textline?

POLIS   Textline

☐   ☐   daily

☐   ☐   more than once a week

☐   ☐   weekly

☐   ☐   fortnightly

☐   ☐   monthly

☐   ☐   irregularly

☐   ☐   only occasionally

4.   Typically, how long do you spend on a POLIS/Textline search?

POLIS   Textline

☐   ☐   1-5 minutes

☐   ☐   6-10 minutes

☐   ☐   11-20 minutes

☐   ☐   21-30 minutes

☐   ☐   various enormously

5.   In the Branch Library, you search POLIS in a menu-driven, simplified form. There is also a more complicated, command driven version of POLIS, which enables you to formulate your own search and get at the data more directly and more quickly.

Would you be interested in using and being trained in the use of the command driven version of POLIS? [Training would take one day.]

☐   yes

☐   no, not really

6.   If you use POLIS, how frequently do you conduct the following types of searches

(a)   a search for information on a subject/topic, not associated with an MP

☐   very frequently

☐   frequently

☐   occasionally

☐

seldom

(b)   a search for information on a subject/topic in connection with an MP

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(c)   a search for all mentions of a given MP

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

7.   If you use POLIS, how frequently do you use it for the following types of searches?

(a)   parliamentary questions

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(b)   progress of legislation (UK and EC)

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(c)   other parliamentary proceedings and papers

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(d)   information about committees, working parties, etc.

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(e)   library material (including official publications, books and deposited papers

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

8.   If you use Textline, how frequently do you conduct the following types of searches?

(a)   a company search

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(b)   a search on business/economics information

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(c)   a search on social/political information

☐   very frequently

☐   frequently

☐   occasionally

☐   seldom

(d)   a search on EEC information

☐	very frequently
☐	frequently
☐	occasionally
☐	seldom

9.   Have you experienced any problems when using POLIS/Textline?

POLIS	Textline	
☐	☐	yes, frequently
☐	☐	yes, occasionally
☐	☐	no
☐	☐	can't remember

If you answered 'yes' to the above question, were these difficulties associated with

		POLIS	Textline
(a)	getting access to a terminal	☐	☐
(b)	equipment failure	☐	☐
(c)	understanding the menu	☐	☐
(d)	formulating a search	☐	☐
(e)	printing the data	☐	☐

10.   Are you satisfied with the results you get from POLIS/Textline?

POLIS	Textline	
☐	☐	yes, usually
☐	☐	sometimes
☐	☐	not normally

11.   If you are a user of both on-line systems, which do you find

(a)	more useful	☐	POLIS	☐	Textline
(b)	easier to use	☐	POLIS	☐	Textline
(c)	cannot compare	☐			

12. Would you prefer the library staff to do the on-line searching for you?

☐   yes

☐   no

☐   I have no preference

13. Would you like to have on-line searching facilities in your office?

☐   yes

☐   no

14. Do you use on-line searching facilities elsewhere?

☐   yes

☐   no

If 'yes', please specify what system and where:

....

....

15. If you have any suggestions as to how on-line searching at the House of Commons can be improved, please feel free to write your comments below.

....

....

....

....

....

THANK YOU FOR YOUR CO-OPERATION

David Nicholas
Polytechnic of North London
School of Information Studies
Ladbroke House
62-66 Highbury Grove
London N5 2AD

Tel.: 01-607 2789 ext. 5078

APPENDIX 3

INTERVIEW SCHEDULE FOR HOC LIBRARY STAFF

Background
Job, department, indexing unit?

Use of POLIS
How long searching. Frequency; types of search; controlled/default; menu/command; search facilities used; attitudes towards; best for.
Observation on MPs and Research Assistants use or non-use.

Use of TEXTLINE
How long searching. Frequency; types of search; controlled/default; menu/command; search facilities used; attitudes towards; best for.
Observation on MPs and Research Assistants use or non-use.

Use of other on-line systems.
Comparisons

APPENDIX 4

ORIEL ROOM OBSERVATION SCHEDULE

Nature of dialogue between MPs and Library staff.

System searched

Files searched

Keywords and approaches used to search databases

Success or otherwise of the search

Clarification of the above also obtained by follow-up interview

APPENDIX 5

THE POLIS RECORD

1.	ACCESSION NUMBER	381180308
2.	GROUP	PQ
3.	TYPE	PNQ
4.	DATE	18:03:87
5.	REFERENCE	112 C927-8
6.	SESSION	86/87
7.	CORPORATE AUTHOR	Dept of Transport
8.	MEMBER	Snape, Peter; Mitchell, David
9.	DESCRIPTION	Statement about the 1,600 redundancies announced at British Rail Engineering Ltd. at Crewe, Derby, Doncaster & York. Supplementary that railway electrification programme & placing of orders for rolling stock with British companies would improve prospects with BREL.
10.	SUBJECT INDEX	Redundancy; Rolling stock; Repairs & maintenance; Railway electrification
11.	ORGANISATION	British Rail Engineering

1.   ACCESSION NUMBER: Control information only.
2.   GROUP: indicates the broad class publication form of the document; in this case, a Parliamentary Question.
3.   TYPE: denotes the specific form of the document presented; in this case, a Private Notice Question (House of Commons).
4.   DATE: Date of PQ, debate, etc. In the case of item types CM, HC, UP, PB, SI and BILL it indicates the date laid before the House.
4A.   STATS INDICATOR: If For an item containing significant statistical information, an extra line appears at this point, giving a note of the kind of statistics contained in the document.
5.   REFERENCE: for Questions, a reference similar to 112 c927-8 refers to Hansard (Commons or Lords as may be appropriate) volume 112, columns 927-8. W refers to the sequence of written Answers at the end of each daily Commons Hansard. For types other than WPQ, OPQ, LPQ, PNQ, CH, BK and OP, the source will generally be the number of the paper concerned.
6.   SESSION: the Parliamentary Session into which the date in section 4 fell.
7.   CORPORATE AUTHOR: this is the name of the Department answering (in the case of Questions) or the body or committee originating the document.
8.   MEMBER: in the case of Questions and Adjournment Debates this gives details of (first) the name of the Member asking, and (second) of the Minister replying. In the case of Early Day Motions, the one name given is that of the Member originating the motion. In the case of debates, up to four names may be employed.
9.   DESCRIPTION: This includes a description of the subject of the Question or the title of the paper. In the case of Questions, a summary of the substance is given, and where possible, an indication of the answer. The full text is not necessarily given and enquirers are referred to the appropriate Hansard for a complete reference to the question and answer.
10.   SUBJECT INDEX: words/terms taken from the HC thesaurus to index the document.
11.   ORGANISATION: the organisation(s) subject of the document.

Source: based upon the P10 document POLIS Prints (1987).

APPENDIX 6

INTERVIEW SCHEDULE FOR JOURNALISTS

Open-ended questions

Areas of questioning

1. Nature of job and typical working week (Diary entries).

2. General information needs: subject, nature, authority, viewpoint, currency, speed of delivery, nationality, processing/packaging.

3. Information seeking behaviour
Constraints on information seeking.

4. Sources of information

5. Use of information systems

6. Major on use, non-use and general awareness of on-line systems and computerised information services. Attitudes towards.

7. Specific questioning on FT PROFILE awareness, use, searching styles, image.

8. Others used.

APPENDIX 7

OBSERVATION SCHEDULE FOR JOURNALISTS

Areas of observation

1. Manner in dealing with information

2. What information is used/on the desk.

3. Sources of information - telephone, wires etc.

4. Any interaction with information systems - description of this.

5. Interactions with library staff

6. Constraints - time, period of time for digestion of information.

APPENDIX 8

INTERVIEW SCHEDULE FOR GUARDIAN LIBRARIANS

BROAD AREAS OF QUESTIONING

Own use of PROFILE
Types of searches, types of enquiries, search commands and facilities used, attitudes towards.

Journalists' use of PROFILE

Observations on searching style, success/failure, volume of use, training, examples, attitudes towards, active departments/individuals.

APPENDIX 9

INFORMATION NEEDS INTERVIEW SCHEDULE
FOR GUARDIAN JOURNALISTS

Open-ended questions

Areas of questioning

1. Nature of job and typical working week (Diary entries).

2. General information needs: subject, nature, authority, viewpoint, currency, speed of delivery, nationality, processing/packaging.

3. Information seeking behaviour
Constraints on information seeking.

4. Sources of information

5. Use of information systems

6. Major on use, non-use and general awareness of on-line systems and computerised information services. Attitudes towards. Changes

7. Specific questioning on FT PROFILE.

APPENDIX 10

QUESTIONNAIRE FOR GUARDIAN JOURNALISTS

We are investigating the possibilities of providing you with better facilities for on-line searching at your desk..

Ideally we would like your name, but if you wish to remain anonymous can you please state whether you are a reporter or a sub-editor and what department you work for.

1.   **How often do you use PROFILE (World Reporter)?**

----

2.   **At what time do you tend to search?**

(you may ring more than one block)

10am-2pm   2pm-6pm   6pm-8pm   8pm onward

3.   **How long do your searches take on average?**

1-10 mins   11-20 mins   more (specify) ....

4.   **When you search PROFILE which files do you search regularly:**

- Guardian   YES/NO

- UKNEWS as a group   YES/NO

- Other papers specifically: YES/NO

     Indy   YES/NO

     Times   YES/NO

     Teleg   YES/NO

     FT   YES/NO

     Others (specify)   YES/NO

- Wires   YES/NO

- Other sources   YES/NO

5. **When you search, what kind of information are you looking for?**

- dates of articles   : 1 2 3 4 5 6 7 8 9 10

- specific by-lines   : 1 2 3 4 5 6 7 8 9 10

- spelling checks   : 1 2 3 4 5 6 7 8 9 10

- background story check : 1 2 3 4 5 6 7 8 9 10

- other (specify)   : ....

6. **How far back do you search on average?**

1-6 months;   6-12 months;   12-24 months;   more (specify) ....

7. **Do you use commas in your searching?**   YES/NO

Do you use + signs?   YES/NO

Do you use brackets in your searching   YES/NO

Do you save your search?   YES/NO

Once you have logged off, your search is retained.
Do you know how to access this?   YES/NO

Can you write down the search term for:

- searching by by-line ....

- searching by date ....

- truncating ....

8. **How often are you completely successful in your search?**

(i.e. you find the information you want)

- nearly always

- sometimes

- rarely

9. **When you want to use PROFILE do you:**

1 - always do your own searches?

268

2 - sometimes ask the librarians to search for you?

3 - only ask the librarians to search for you?

10. **If you ticked 2, why do you ask the library?**

- can't get through to PROFILE

- have not the time

- it is a difficult search

- think librarians are better searchers

- other (specify) ....

11. **Do you start an on-line search before checking out cuttings files?**

- nearly always

- sometimes

- rarely

12. **Would it be useful to have an index of lib. files on Atex?**   YES/NO

Would you still start a search without checking files?   YES/NO

13. **How willing are you to delegate on-line research to the library?**

- very willing

- not very willing

- if not very willing, please state why ....

14. **How willing are you to be trained on PROFILE?**

- very willing

- not very willing

The Guardian is now available on CD-ROM which we might consider networking around the building.

15. **How important are images of cuttings to you?**

- very important

- quite important

- not very important

* If you answered 'very important' please state why ....

....

16.   How important is it to you to have access to files other than The Guardian?

- very important

- quite important

- not very important

Any Comments?

Thank you very much for your co-operation. Please state your name and/or whether you are a reporter or sub-editor, your department and return this questionnaire to the library by October 31st.

SAMPLE PAGE FROM FT PROFILE COMPUTER LOG

22FEB95:09:45:53	AMYG	**LIBRARY**
22FEB95:09:46:07	AMYG	**INTNEWS**
22FEB95:09:46:08	AMYG	**SWB,RTN,AP,KEE**
22FEB95:09:46:19	AMYG	**SEL SWB,AP,RTN,**
22FEB95:09:46:37	AMYG	*G SEA//POLLUTY **
22FEB95:09:46:41	AMYG	*PK SEA@ *12START*
22FEB95:09:46:45	AMYG	SB
22FEB95:09:46:48	AMYG	*PK SW EA*1@START*
22FEB95:09:47:05	AMYG	*G SEA*1//POLLUT**
22FEB95:09:47:18	AMYG	*G (SEA*1.. //POLLUT*)@START*
22FEB95:09:47:25	AMYG	M
22FEB95:09:47:39	AMYG	M
22FEB95:09:47:56	AMYG	H ALL
22FEB95:09:48:29	AMYG	X
22FEB95:09:57:50	AMYG	*PK BRITAIN,ITALY,GERMANY MANY,RUSSIA*1,IELANDCE*
22FEB95:09:57:51	AMYG	
22FEB95:09:58:00	AMYG	H ALL
22FEB95:09:58:20	AMYG	X
22FEB95:09:58:22	AMYG	CTX ALL
22FEB95:09:58:30	AMYG	
22FEB95:09:58:39	AMYG	
22FEB95:09:58:49	AMYG	
22FEB95:09:59:02	AMYG	X
22FEB95:09:59:06	AMYG	H 6-
22FEB95:09:59:17	AMYG	X
22FEB95:09:59:19	AMYG	TX 13
22FEB95:09:59:30	AMYG	
22FEB95:09:59:39	AMYG	H 10-
22FEB95:09:59:51	AMYG	X
22FEB95:09:59:53	AMYG	TX 16
22FEB95:10:00:04	AMYG	
22FEB95:10:00:12	AMYG	H 20-
22FEB95:10:00:28	AMYG	X
22FEB95:10:00:29	AMYG	TX 26
22FEB95:10:00:38	AMYG	X
22FEB95:10:00:42	AMYG	H 27-
22FEB95:10:00:54	AMYG	
22FEB95:10:01:14	AMYG	
22FEB95:10:01:30	AMYG	
22FEB95:10:01:41	AMYG	
22FEB95:10:01:58	AMYG	
22FEB95:10:02:09	AMYG	
22FEB95:10:02:15	AMYG	X
22FEB95:10:02:17	AMYG	TX 88
22FEB95:10:02:35	AMYG	X
22FEB95:10:02:37	AMYG	H 89-
22FEB95:10:02:48	AMYG	X
22FEB95:10:02:49	AMYG	**END**

DATA COLLECTION FORM FOR EXTRACTING DATA FROM PROFILE LOGS

SESSIONS	
TIME	
SEARCHES	
TIME	
TERMS (phrases inc)	
STAGES	
COMMAND	
(SEARCH)	
(SUBJECT)	
GET	
PICK	
* *(TRUNCATION)*	
(BOOLEAN)	
+ *(AND)*	
, *(OR)*	
- *(NOT)*	
() (NESTING)	
(PROXIMITY)	
(PHRASE)	
/ (SENTENCE)	
// (PARAGRAPH)	
(FIELD)	
@START	
@HL	
@BL	
GET DATE	
PK DATE	
FILE SELECTION	
NUMBER	
uknews	
GDN	
Other files	
(OTHER CMNDS)	

SAVE/RUN	
STEPBACK	
HIGHLIGHT	
M (MORE)	1m 2m

(DISPLAY CMNDS)	
H (HEADLINE)	
CTX (CONTEXT)	
TX (FULL-TEXT)	
NUMBER (per srch)	
ALL	
SCREENS	
no hits	
RESULTS IN TX	
NOBREAK	

ERRORS	
INPUT/SPELLING	
WRONG . CMD.	
MISSING CMD.	
.POOR LOGIC	