SDI services in the UK in the field of science and technology

PUBLISHER

Loughborough University of Technology

S.D.I. Services in the U.K. in the field of Science and Technology.

Jean F. Remington

Supervisor - R.A. Wall F.L.A.

Submitted in partial fulfilment of the requirements for a Master's Degree of Loughborough University of Technology.

1970

ACKNOWLEDGEMENTS

The author wishes to thank Mr. R.A. Wall, Deputy Librarian, Loughborough University of Technology for his help and guidance throughout this work; Mr. David Rankin for his advice on the questionnaire; Mrs. Marion Henshall for her assistance with the typing and Mr. Brian Neale for assistance with corrections and encouragement throughout. The author also gratefully acknowledges the cooperation and assistance of S.D.I. systems operators and users who supplied much of the information.

CONTENTS

SUMMARY.

Mechanised S.D.I. systems in the U.K. are surveyed, together with some associated narrow subject current awareness services. Techniques and procedures used in S.D.I. systems are discussed in detail. The place of S.D.I. in an individual's information system and also in the national information network of this country, is considered.

ABBREVIATIONS

A.C.S. - American Chemical Society.

A.S.C.A. - Automatic Science Citation Alert.

C.A. - Chemical Abstracts.

C.A.C. - Chemical Abstracts Condensates.

C.A.S. - Chemical Abstracts Service.

C.B.A.C. - Chemical Biological Activities.

C.C. - Current Contents.

C.C.P. - Current Chemical Papers.

C.P.P. - Current Papers in Physics.

C.S.R.U. - Chemical Society Research Unit.

C.T. - Chemical Titles.

E.I.U. - Experimental Information Unit.

E.L.D.O. - European Launcher Development Organisation.

E.S.R.O. - European Space Research Organisation.

G.K.N. - Guest Keen and Nettlefolds.

I.A.A. - International Aerospace Abstracts.

I.B.M. - International Business Machines.

INDATA - Instrument Data.

INSPEC - Information Services in Physics, Electrotechnology and Control.

I.S.I. - Institute of Scientific Information.

I.S.I.P. - Iron and Steel Institute Profiles.

L.S.E. - Liason Scientist Experiment.

MEDLARS - Medical Literature Analysis and Retrieval System.

N.A.S.A. - National Aeronautics and Space Administration.

N.B.S. - National Bibliographic Service.

N.L.L. - National Lending Library for Science and Technology.

N.S.A. - Nuclear Science Abstracts.

O.S.T.I. - Office of Sientific and Technical Information.

POST - Polymer Science and Technology.

R.A. - Research Association.

SCAN - Selective Current Awareness Notes.

S.C.I.P. - Students Chemical Information Project.

S.D.C. - Scientific Documentation Centre.

S.D.I. Selective Dissemination of Information.

S.D.S. - Scientific Documentation Service.

S.I.C. - Specialised Information Centre.

S.R.C. - Science Research Council.

S.T.A.C. - Scientific and Technical Aerospace Reports.

S.T.L. - Standard Telecommunications Laboratories.

U.K.C.I.S. - United Kingdom Chemical Information Service.

Journal Abbreviations

Amer. Doc. - American Documentation.

Aslib Proc. - Aslib Proceedings.

Inf. Sci. - Information Scientist.

J. Doc. - Journal of Documentation.

J. Chem. Doc. - Journal of Chemical Documentation.

CHAPTER 1

INTRODUCTION

This dissertation is a report on the present state of selective dissemination of information in the field of science and technology, in this country.

H.P.Luhn (64) who first coined the phrase 'selective dissemination of information', described it as "that service within an organisation which concerns itself with a channelling of new items of information , from whatever source to those points within the organisation where the probability of usefulness , in connection with current work or interests , is high."

S.D.I. is taken in this dissertation to be - a personalised current awareness service whereby 'profiles' of user's interests are matched against incoming documents and users are notified of documents which are likely to be of interest to them. This matching can be done manually or by computer, but S.D.I. is usually taken to mean mechanised matching. Here both forms of S.D.I. will be considered to some extent. However, this dissertation is mainly concerned with mechanised services, particularly 'custom'S.D.I. 'In house' S.D.I. systems run by organisations for their own staff and 'quasi'S.D.I. services where users choose to receive output from standard profiles or subject categories, are also considered.

The basic ingredients of mechanised S.D.I. systems are diseussed in detail in Chapter 3.in an attempt to deduce the best procedures for an S.D.I. system in different circumstances.

Most organisations, where the value of information of one sort or another is appreciated, have developed their own types of information systems. This is often quite an unconcious act between personnel working together It may be simply a librarian or information officer channelling information by means of a bulletin or by individual notifications. This latter method is often practised in science and technology and apparently works well. But the time comes when one information officer cannot scan all the incoming literature and the number of users one worker can serve is limited. In many organisations the point has been reached when some new method of information dissemination is required. The question is whether S.D.I. is the answer.

An attempt was made to discover what form of information dissemination was practised amongst research workers in different environments. In Chapter 4 the information dissemination methods available to different types of research worker are disscussed and conclusions are reached as to the best type of information dissemination systems for research workers in different environments and the place within these of S.D.I. in its various forms. The relative importance of different types of S.D.I. service to research workers in different environments together with the most suitable role for custom S.D.I.

This dissertation constitutes a survey,of S.D.I. systems in this country, an analysis of S.D.I. systems in general and a discussion of their importance to individual users and in the information network of this country.

CHAPTER 2.

SURVEY OF U.K. S.D.I. SYSTEMS

In this chapter existing and past mechanised S.D.I. systems operating in this country are surveyed together wibb a few closely related current awareness services which are considered to be 'quasi' S.D.I. services. The systems are described by breaking them down into their component parts rather as Bivona and Goldblum (12) did in their survey of U.S.A. S.D.I. systems, that is as follows:-

1. Backround ofsystems.
2. Coverage of systems.
   Subject coverage.
   Form coverage.
   Foreign coverage.
3. Input for system.
   Input method.
   Input information.
   Input volume.
4. Profile construction and match strategy.
   Methods of profile construction and modification.
   Profile form and match criterion.
5. Output from system.
   Outout frequency and uptodateness.
   Output form.
   Output information.
   Other outputs from system for users.
6. Users of system.
7. Statistics and evaluations of systems.
8.C Costing of systems.
9. Computing of system .
   Computing time,Hard ware and Software

A.S.C.A. - Automatic Science Citation Alert.

Institute ofScientific Information.

U.K. agent - A.E. Cawkell,

132, High Street,

Uxbridge.

1. Background.

The A.S.C.A. service is a well established commercial service run by I,S,I, The A.S.C.A. service is based on the magnetic tapes used to produce Science Citation Index. In 1965 A.S.C.A. I was launched. This service could only pick out references which cited named papers. The A.S.C.A. II service which followed could search on source authors, source organisations and patent classifications and assignees, A.S.C.A. III was introduced at the end of 1966 when the Permuterm Subject index facility was introduced to S.C.I. and this enables searches on words in title as well. With increased coverage and source organisation definition down to departmental level A.S.C.A. IV is the service of 1969.

2. Coverage.

Subject coverage:- The service aims to be multidisciplinary, covering natural sciences, physical sciences, medecine and technology.Its scope steadily extends to new subjects.

Form coverage:- 2,200 journals are completely covered (including editorials, letters, corrections and citations etc.). Citations in articles bring in book, report, patent, and conference material.

Foreign language material is translated into American, cover to cover translations are used where possible.

3. <u>Input.</u>

Method:- Search on tapes already prepared in the production of S.C.I.

Information:- Authors(up to 10), titles(in English), journal titles(S.C.I.abbreviation), vol.nos.,iss.no., starting page no.,year,code for type of term,organisation, no. of cited refs and details of these, journal accession no.

Volume:- 390,000 source items per.year (Aug.1968).

4 .<u>Profile construction and match strategy</u>

Profiles are constructed by users on supplied forms

Profiles may be modified as desired. Modification is immediate and free

Profiles consist of words, word fragments or phrases, cited references, cited authors(first author only), source authors, source organisation, and source journal titles taken alone or in combinations and linked by Boolean logic

5. <u>Output.</u>

Weekly. Eighteen days minimum elapse between the publication of the primary journal and S.D.I. notification from it. Output is on continuous sheet or for several users in the same organisation it can be on magnetic tape in user blocks. Printout gives details of which search terms caused the hit. Details include author,title, journal reference with year and number of citations.

Users are periodically provided with cumulated statistical tabulations of their weekly A.S.C.A. reports.

6. Users.

Approximately 50 users in the U.K. These include the S.I.C. in Intestinal Absorbtion at Sheffield University, National Physical Laboratory(24 profiles) and several industrial organisations (not only those with chemical interests.).

8. Costing.

Costing is on the basis of expected output quantity These are based on the search term types chosen, terms taken in combination are cheaper.

Term type	Cost p.a.for single term $	combination term $
words	6.00	4.00
cited ref.	2.00	
cited author	6.00	
source author	4.00	
organisation source	10.00	
source journal	6.00	4.00

High frequency words and source journal terms which are likely to produce more than the average number of references are priced individually above the normal price for that term type.

Comments.

A.S.C.A.'s multidisciplinary approach has made it attractive tousers whose interests were not catered for by C.S.R.U. It is to be hoped that soon S.D.I. services will be available in the U.K. in a wide variety of subjects. A.S.C.A.'s coverage is best in chemistry and biology. The citation facility of A.S.C.A. will always make It important for interdisciplinary subjects and other difficult subjects. However, A.S.C.A. is rather expensive. L.S.E.profiles cost$383

References.

CAWKELL A.E. Private communications.

CAWKELL A.E. Search strategies using the Science Citation Index. in Computer Based Information Retrieval Systems. ed. B.Hougton 1968 Clive Bingley.

GARFIELD E., SHER I.H. A.S.C.A.- A new personalised current awareness service for scientists. In The American Behavioral Scientist. 1967 X (5) 29-32.

GARFIELD E., SHER I.H. I.S.I.'s experience with A.S.C.A. - a selective dissemination system. In J. Chem.Doc. 1967 7 (3) 147- 153.

E.S.R.O./E.L.D.O. Space Documentation Service.

114, Avenue de Neuilly,

92 - Neuilly,

France.

1. Background.

Although this service can hardly be called a U.K. service, it is the European service in this field and so requires mention as there are several U.K. users of the service. The Space Documentation Service (S.D.S.) was set up by the European Space Research Organisation (E.S.R.O.) and the European Launcher Development Organisation (E.L.D.O.) in collaboration with EUROSPACE, as part of an overall scientific and technical information programme to make available the results of space research. The basis of the system is an exchange agreement between N.A.S.A. and E.S.R.O./E.L.D.O. which has been operational since 1962.

S.D.S. services include, bibliographic search facilities, a reproduction service for document supply to authorised users, a standard profile service and databank(these last two services are descibed in section 3 of this chapter) and an S.D.I. service described below.

2. Coverage.

Space sciences.

Scientific and Technical Aerospace Reports (S.T.A.R.) published by N.A.S.A. which includes:-

Scientific and technical reports by N.A.S.A. and its contractors.

Scientific and technical reports of government agencies, universities and research organisations throughout the world.

Scientific and technical articles prepared by N.A.S.A.

employees that are published in learned and technical journals.

International Aerospace Abstracts (I.A.A.) published by the American Institute for Aeronautics and Astronautics in collaboration with N.A.S.A. This includes the contents of new scientific books and approximately 1,400 periodicals.

3. Input.

Magnetic tapes of S.T.A.R. and I.A.A. provide the input.

These contain indicative abstracts and details of subjects, authors, originating body and source references. all items are indexed by 10-20 keywords from the N.A.S.A. thesaurus.

Input amounts to 6,000 to 7,000 items monthly.

4. Profile construction and match strategy.

Profiles are constructed by analysts at the headquatres in France. This done from users statements which should contain:- natural language desciptions of interests, terminology explanations and synonyms, concepts not wanted, concepts of particular interest,(these should be in N.A.S.A. thesaurus allowed terms), S.T.A.R. subject category, N.A.S.A. thesaurus subject category, and a list of document references directly relevant.

There is a facility for users to see their profiles before they are run. Any modification of profiles will be at the users instigation as no user feedback is used. RECON the new remote console means that profile modification can be immediate.

Profiles are in the form of N.A.S.A. thesaurus terms linked by Boolean logic or group weighting. Matching is done on the E.S.R.O. computer Darmstadt, Germany.

5. Output.

Monthly, is about three to four months after the appearance of the item in its original publication, though this varies considerably.

Information output includes:- S.T.A.R./I.A.A. document no., N.A.S.A. subject category no., author, title, corporate source, date of publication, pagination, if any references are cited, language, source reference, keywords allocated with those appearing in hardcopy indexes marked, and an indication of whether the item is available from S.D.S. on micro fiche.

8. Costing.

Approximately £33 per profile.

Comments.

Here is a fairly cheap service in the applied sciences which covers a wide range of subjects. It would probably provide an attractive service to many users in this country if they knew of it.

References.

RAITT D.I. Private communication.

INSPEC (Information Service in Physics, Electrotechnology and Control) S.D.I. Investigation.

Institute of Electrical Engineers,
26, Park Place,
Stevenage.

Manager - P.Clague.

1. Background

The INSPEC S.D.I. service is a purely experimental service which is part of the larger INSPEC project, concerning the computerised production of Science Abstracts and the three Current Papers publications. This S.D.I. investigation is a continuation of the National Electronics Research Council project started in 1965. This project is also supported by an O.S.T.I. grant. The aim of the project is to test the feasibility of S.D.I. by determining the value, economics, efficiency and acceptability of an S.D.I. system.

Draft profiles were devised and run experimentally for six months. From November 1968 users have been receiving a regular service which is expected to run for a year. This is to be continued on a cost recovery basis in 1970. Meanwhile the full INSPEC data base tapes will be available from January 1970 and program are being devised, with advice from U.K.C.I.S., for an S.D.I. service based on these full tapes.

2. Coverage.

Subject:- Electronics research.

Forms:- 300 journals from Science Abstracts coverage are covered selectively. These include English language articles only though these may come from cover to cover translations foreign journals with English language articles.

3. Input.

Method:-Details are keypunched. Eventually the full INSPEC tapes will probably form the data base.

Information:- Article no.,broad subject groups code, authors, titles, journal reference(jnl. title abbreviation, vol.no., iss. no., inclusive pages, date of issue),and descriptors(average 12,including author name and journal coden). Only the descriptor block is searchable.

Volume:- Input is expected to be 240 items per week.

4. Profile construction and match strategy.

Profile construction is being experimented with. The basic method being used is for users to describe their intecests in a narrative form, explaining terminology and supplying six very relevant references. From these details profiles were drafted.

Profiles are modified using user feedback. Users rate each meference they receive witb:- 1 - highly relevant.
2 - of some relevance.
X - not relevant.

Profiles consist of descriptors linked by Boolean logic. These descriptors can be authors names or journal codens though the majority are thesaurus terms. The thesaurus is based largely on that compiled in connection with the work to revise the indexes for Electrical and Electronics Abstracts and is similar to the Engineering Joint Council Thesaurus.

5. Output.

This is also being experimented with. Users receive either cards bearing author, title, journal reference and alist of descriptoos of individual documents or continuous computer printout with slightly less detail. The possibility of supplying abstracts will also be experimented with. It is hoped that only seven or, eight days will elapse between articles entering the system and notifications being dispatched.

6. Users.

540 statistically chosen electronics research workers in the U.K. and 60 others who are part of group profiles. These research workers are in a variety of enviromments.

7. Statistics.

Statistics are to be collected on:-

Relevance as measured by user ratings of each article.

Recall by marking up of the total systems output by users.

The effect on these averages of different methods of profile construction.

The effect on acceptibility to users of different output formats and other variables.

The effect on users of the service is to be measured by questionnaire and by observation by users librarians, diary techniques and some interviewing.

9. Computing.

Programs were devised by English Electric for the KDF 9 computer and are run by the Document Processing Centre of O.S.T.I. on National Computer Centre computers in Manchester

Comments.

For S.D.I. to be of much use as with any other information service it must be designed to suit the user. It is to be hoped that this experiment will reveal the true value of an S.D.I. service.and that INSPEC will find it worth while continuing the service. on the full INSPEC tape data base. This latter is not likely to happen till 1971.

Users of this experimental service seemed to have little idea how the system worked. Their initial statements of interests were obtained from them before they had any understanding of what the system could provide for them, thus these would tend not to produce good profiles. The profile modification method did not appear to work very well either. Non of the users intervied seemed to be getting a satisfactory service (see Appendix II). With better user education and a broader data base a more satisfactory service could have been supplied to the users interviewed. It will be interesting to see if users will be prepared to pay for this service.

References.

AITCHISON T.M. et.al. Developments towards a computer based information service in Physics, Electrotechnology and Control. in Information Storage and Retrieval. 1968 4 (2) 177-186.

AITCHISON T.M. Investigating S.D.I. in Aslib Proc. 1965 17 (12) 343-350.

BLEE M. INSPEC project. in Data Systems April 1968 38-40.

CLAGUE P. Private communication.

CLAGUE P. The S.D.I. investigation in Electronics in Computer Based Information Retrieval Systems. ed. B.Hougton 1968 Clive Bingley.

EAST H. Design of a computer aided information service for Physicists with minimum concessions to the machinery. in Inf. Sci. 1967 1 (3) 107-119.

MARTIN M.D.,SMITH J.R. INSPEC - a new concept in information services. in Electronics and Power Feb. 1969 15 66-69.

MOUNTBATTEN,EARL Controlling the information explosion in The Radio and Electronic Engineer 1966 31 (4) 195-208.

MEDLARS (Medical Literature Analysis and Retrieval System).

U.K. agent - Dr. A J. Harley

N.L.L.,

Boston Spa,

Yorkshire.

1. Background.

Since 1963 Index Medicus which is produced by the National Library of Medicine in U.S.A.,has been computer printed. The magnetic tapes from this production, that is the input tapes minus the layout instructions, have been used for computer retrospective searching in lieu of lengthy scanning of Index Medicus. Searches in this country have been dealt with by Dr. Harley at the N.L.L. with the aid of the Newcastle University KDF 9 computer. Proceedures having been well established for retrospective searching are now being adapted to provide an S.D.I. service which at present is in a very experimental state.

2. Coverage.

Subject - That of Index Medicus, a comprehensive cover over the field of medicine.

Forms - Journals only are covered. Approximately 2,300 are covered but some only selectively.

3. Input.

Method - Magnetic tapes from the publication of Index Medicus.

Information - Titles, authors, journal title abbreviation (Index Medicus abbreviation), vol. no., inclusive pagination, date of issue, language of article, list of subject headings (up to 30, the average no. is 10), a form code(biography etc.)

Foriegn language titles follow their translated version.

4. Profile construction and match strategy.

Profiles are constructed by users and search editors at N.L.L.

Users provide feedback by rating each notification as:- important, relevant, don't know, or irrelevant. Users are also asked to list critical references, these are particularly important references which are notified, and important references missed by the system but found elsewhere.

Profile questions are usually divided up into three subsearches at different levels of specificity. Terms taken from MESH (Medical Subject Headings) are linked by Boolean logic to form the user profiles.

5. Output.

Monthly, on average six months after the publication of the item originally.

The computer printout gives details .- author(s), title in American and original language, journal eeference and list of index terms assigned to the article with matched terms marked.

6. Users.

At present 80 profiles are being run as part of the S.D.I. experiment. 25 profiles are being run as part of the Liason Sientist Experiment.

7. Statistics.

Statistics on theS.D.I. experiment are expected to be available at the beginning of 1970.

9. Computing.

Approximately 2½ hours computing time is required for 20 - 30 questions for retrospective searching.

A KDF 9 computer is used for processing.

Comments.

Being based on a thesaurus MEDLARS has quite a few disadvantages. Index Medicus use is made easier by its thesaurus base and simple searches may not warrant the use of MEDLARS . It has been found that 30% of retrospective searches on Index Medicus would produce all references found in the MEDLARS search. Athesaurus restricts the subject fields for which the service can be used. Ther are many fields peripheral to medicine where MEDLARS might be thought to be of use. But with limiting thesaurus language designed for other users it is near impossible to express interests precisely. The MEDLARS evaluation report (59) showed that the performance of the system was worst for topics in the Behavioral sciences and Drug/Biology fields. Difficulties were expressed by workers in these fields who were interviewed in the user survey in Appendix II.

Medlars can hardly be called a current awareness service being on average six months out of date and a minimum of three months late. One Pharmacologist found the service useful as a check on his other current awareness methods.

References.

HARLEY A.J. Private communication.

HARLEY A.J. MEDLARS. A comparison with hand searching in Index Medicus. in Inf. Sci. 1968 2 (2) 59-72.

HARLEY A.J. MEDLARS in context. in J. of Royal College of Physicians of London 1967 1 (3) 261-270.

HARLEY A.J. The Medical Literature Analysis and Retrieval System - MEDLARS. in Computer Based Information Retrieval Systems. ed. B.Hougton 1968 Clive Bingley.

HARLEY A.J. U.K. MEDLARS Information Retrieval Handbook for Users. N.L.L.

LANCASTER F.W. Evaluation ofthe MEDLARS demand search service U.S. National Library of Medicine . Jan. 1968 278p.PB 178 660.

Plasma Physics Current Awareness Service.

U.K.A.E.A. Culham Laboratory,
Culham,
Berkshire. Information officer A.G. Cheney.

1. Background.

In 1964 after three years of operating a manual current awareness service to internal users, shear bulk began to make the service impossible. With the availability of an English Electric KDF 9 computer experiments in a computerised current awareness service began. The service operates at two levels, providing a library bulletin and an S.D.I. service for individuals from the same input. The system became operational in April 1966. In 1967 O.S.T.I. offered a grant for two years for the service to be extended to external users experimentally. Now the O.S.T.I. grant is on the point of running out, the S.D.I. service to external users is to be continued on a cost recovery basis, probably from Easter 1970. The library bulletin previously widely distributed gratis is to become a commercial publication. The service is mainly run by three people with the duties divided up into, Scanning for input, output scanning and user service, and general coordination and internal user liason.

2. Coverage.

Subject - Plasma Physics and related subjects including some lasers.

Forms - items for input are selected from the library acquisitions of over 600 journal titles. 1,300 reports, 700 books, and 250 pamphlets per annum.

3. Input.

Method:- Details are keypunched on a paper tape typewriter.

Information:- For journals, journal code no., vol. no., iss. no., date, title, authors, starting page and code indicating whether for S.D.I. or just the bulletin. For reports, report no. organisation of origin, title, author, date, pagination, classification coding and descriptors taken from the major word glossary.

Volume:- At the beginning of 1968 approximately 75 journal articles and 35 reports were input weekly, with roughly two thirds of the input being used for S.D.I.

4. Profile construction and match strategy.

Profiles are constructed by the information staff. In the case of external users this is based on user statements which include details of:- subjects of interest, authors of importance, aspects of the subject not of interest, three relevant recent articles or reports, any items authored by the user recently and details of other current awareness services received.

Profiles are modified in the light of the relevance of output. Printouts are scanned by information staff who check for obvious irrelevancies and by user feedback. User relevance rating is on the basis of:-a)relevant to subject interests.

b) of interest but not directly rel.

c) irrelevant.

d) if relevant, whether item was known of previously.

Profile construction is quite complicated and time consuming. Thus frequent profile modification is not actively encouraged as at present this necessitates the whole profile tape being reset. With new hardware on-line modifications willbe possible

Input is arranged into set data blocks and major words are separated from titles by comparison with the major word glossary, which is built up from the major words in the profiles Profiles define major and ancillary subject interest of users in the form of a matrix. The columns describe the users different interests with the last column reserved for authors names (surnames only and space for 10 per profile). The rows are used in sets of five, twenty five rows. The first five rows are filled with major words and the other twenty with four sets of minor words. Within the groups words are arranged alphabetically for speedy searching. Words are truncated to severn characters, the eighth is a tag. For a match to occur at least one major word must be present linked with minor words in the specified mode. Three linking modes are available specifying different degrees of linking with minor words. A non select mode is also available and diagnostic modes can also be chosen where a printout will be produced for information staff's perusal when only a partial match occurs.

Profiles are matched with items first by author, then by major words and finally minor words.

5. Output.

Weekly. Internal users found around 6% of relevant references were already known to them and external users found around 7% were already known.

Computer printout, upper case only, in filing card size blocks. Details include, journal title, vol. no., iss. no., iss. date, title, author, and starting page.

6. <u>Users.</u>

The estimated possible nomber of internal users is around 300 of whom approximately 90 were receiving S.D.I. at the end of 1967. Plasma Physicists outside Culham from laboratories in America and Europe as well as the U.K. are now receiving Culham S.D.I. notifications. At the beginning of 1968, 85 such people were receiving notifications, half of these overseas. With the new hardware it will be possible to have more users.

7. <u>Statistics.</u>

Relevance measurements were made based on user feedback. The relevance ratio average when a) and b) were used was 90% and when only a) was used it was 85%. These were with some noise removed by information officers scanning of output, though this was declared to have a negligible effect.

Recall measurements were obtained from users marking up the weekly bulletin as , items relevant but not notified and items irrelevant but notified. These measurements were made on a sample basis for the twelve months May 1966 to April 1967. The results obtained were:-

	Recall ratio %	Precision ratio%
Journal articles	76	92
Report titles	56	82

8. <u>Costing.</u>

A comparative costing has been carried out between manual and computer systems for journal input only for bulletin and S.D.I. The results show that the computer system is marginally more expensive, £2,550 to £2,290 though the service provided is superior. More users can be given a service in greater depth with greater uniformity of standard.

9. Computing.

Processing time on the original hardware amounted to 35 minutes per week.

Originally an English Electric KDF 9 computer was used with a core store of 32K and a disc file of 4 million words. Programmes were written in user code by a member of the computing staff who left shortly afterwards hence changes have not been under taken.

A 470 computer is being installed at present and new programmes are being written. An improvement in the service possible is expected.

Comments.

In this specialised subject field some of the advantages of an S.D.I. service are lost when the actual documents are not readily available as is the case for many external users. A document supply service would seem to be desirable here. Because of this difficulty in document supply and in view of the high relevances possible card output would also seem desirable. The advantages of the S.D.I. service over the weekly bulletin is marginal for external users.

References.

ANTHONY L.J.,CARPENTER D.H., CHENEY A.G. S.D.I. Using KDF 9 computer. in Aslib Proc. 1968 30 (1) 40-64.

ANTHONY L.J., CHENEY A.G., WHELAN E.K. Some experiments in the selective dissemination oO information in the field of Plasma Physics. in Information Storage and Retrieval 1968 4 (2) 187-200

CHENEY A.G. Private communication.

United Kingdom Chemical Information Service.(U.K.C.I.S.)
(formerly the Chemical Society Research Unit in Information Dissemination and Retrieval.)

Nottingham University.

Director - Dr. A.K. Kent.

1.Background.

The Unit was set up jointly by the Chemical Society and O.S.T.I. in August 1966 an experimental S.D.I. service was started,for approximately 250 users, based on Chemical Titles (C.T.) and Chemical-Biological Activities (CBAC). The Unit also cooperated with the American Chemical Society (A.C.S.) in the experimental use of these. Later POST-J and POST-P were added to the system. In 1967 O.S.T.I. sponsored the Students Chemical Information Project in connection with the Units(35)(84). Since July 1st 1968 the Unit has been operating a cost recovery service Chemical Abstracts Condensates (C.A.C.) are now being used in an experimental service. The Unit is also beginning to experiment with the use of Biological Abstracts Preview tapes, Food Science and Technology Abstracts and the Index Chemicus Registry System on tape. More profiles are being run paid for by O.S.T.I. for the Liason Scientist Experiment (L.S.E.) which started in 1969. U.K.C.I.S. is cooperating with Inspec in producing programmes for the full Inspec tapes. U.K.C.I.S. is also to work on the exploitation of the C.A.S. file of 3,000 compounds registered by the C.A.S. compound registry system.

2. Coverage.

Custom search services are offered on Chemical Titles (C.T.) Chemical-Biological Activities (CBAC), Polymer Science and

Technology for journals (POST-J) and for patents (POST-P)

A free exoerimental service is being offered on (C.A.C.)

Subject coverage -

C.T. - New work in pure and applied chemistry and chenical engineering.

CBAC - In-depth coverage of topics on the interface between chemistry and biology. The effect of exogenous compounds on biological systems. The metabolism of organic compounds in biological systems. In vitro reactions between compounds of biological interest.

POST- Chemistry, chemical engineering and technology of polymers.

C.A.C.-As for C.A. that is chemistry and related subjects.

Form coverage:-

C.T. - 700 journals of which 200 only are covered completey.

CBAC - 600 journals covered selectively.

POST-P-Current patent literature in the field.

POST-J-Selected current journal articles and government reports in the field.

C.A.C.-12,700 journals,books, and patents.

3. Inout.

Method:- Magnetic tapes received from C.A.S. in the U.S.A. are transcribed from I.B.M. tapes to provide tapes suitable for KDF 9 input.

Information - (brcketed information is non searchable.)

C.T. - Title, authoas, journal coden, (vol. no., iss. no., inclusive pagination.)

CBAC - Title, autnors, journal coden, (vol. no., iss. no., inclusive pagination, year of publication, institution of origin address),digest, C.A.S. registry no., molecular

POST-J - As for CBAC.

POST-P - Title , patent no., assignees, country of origin, digest, inventors.

C.A.C. - Title, authors, journalcoden, vol.no., iss.no., inclusive pagination, language of original publication, keywords (average of 4) , C.A.S. registry no., molecular formula.

Volume:-

C.T. - 125,000 items per year.

CBAC - Approx. 12,500 items per year.

POST-J - Approx. 25,000 items per year.

C.A.C. - Approx. 250,000 items per year.

4. Profile construction and match strategy.

Profiles are constructed by users on supplied forms with assistance from manuals and teachins provided by the Unit.

Profiles may be modified as often as desired. Modification is at the instigation of, the user , is immediate and free.

Users relevance assessments of output were made during the experimental stages.

Profiles consist of search terms linked by weighting and Boolean logic. Search terms can be words, word phrases, word fragments, author names, journal codens, registry numbers or molecular formulae. Item files are matched against the profile file.

5. Output.

Fortnightly. C.T. is only two or three weeks post publication C.A.C. is at least four weeks post publication of the original CBAC is at least ten weeks post publication.

Output is on 6"X 4" cards different colours for different services(with facilities for edge punching). Output can also be in the form of continuous computer printout or on magnetic tape.

Analyses of profile search term productivity is available for users.

6. Users.

There are 170 paying users. As part of the Liason Scientist Experiment (60),(61) 500 profiles are being run on C.T., 75 on CBAC , 20 on POST and a further 70 on C.T. as part of the Evaluation Experiment(61).

7.Statistics.

In the experimental stages a lot of statistics were collected (56).

8. Costing.

The cost of a custom profile is a minimum of £45 which allows for one unit of 50 search terms in the profile and one output unit of 150 items per year. Additional profile search term units can be obtained at £5 each as can additional output units and multiple copies of output (up to 9) at £5 per unit of output per copy.

9. Computing.

Computer time is around 1hr. 20mins for 100 profiles.

The University of Nottingham computer, a KDF 9, is used only by U.K.C.I.S. all day on Saturdays.

Programmes were written by the Unit and were based on the original C.A.S. programmes written for I.B.M. machines.

Comments.

Comparison tests of C.T. and CBAC with A.S.C.A. services by Abbot(1) and Barkla(8),(9) revealed that these services miss references retrieved by A.S.C.A. and vice versa. Analysis of these misses revealed some deficiencies in the procedures.

The limitations of a titles only search such as with C.T. could be overcome by adding keywords as is being done with C.A.C. However, keyword use brings in other problems as is being found with C.A.C. which has increased coverage and greater search depth than C.T. The snag that was revealed in CBAC searches and might also be expected to be found in POST searches was that profile word combinations were occasionally matched with unrelatedwords from different sentences or paragraphs of the CBAC digests. This difficulty could be overcome by searching each sentence or paragraph of the digest as a separate unit. U.K.C.I.S. have solved this problem now.

The selective coverage by C.T. andCBAC means that users in peripheral fields cannot easily supplement the service by just scanning a few added journals.

References.

Chemical Society Research Unit in Information Dissemination and Retrieval.- Search Manual.

KENT A.K. The Chemical Society Research Unit in Information Dissemination and Retrieval. in Svensk Kemish Tidskrift 1968 80 (2) 39-46.

KENT A.K. Problems of remote processing of magnetic tape information files. in Aslib Proc. 1968 20 (11) 502-508.

KENT A.K. Retrieving chemical information by computer. in Chemistry and Industry 7th. Sept. 1968 (36) 124.

KENT A.K. United Kingdom experiences in the operation of a retrieval and dissemination system based on C.A.S. search tapes. Paper presented before the Chemical Literature Division of the American Chemical Society, Atlantic City, Sept. 1968

Guest Keen and Nettlefolds.

G.K.N. Group Technological Centre,

Birmingham New Road,,

Wolverhampton.

Information offieer - Miss. M.E. Fisher.

1. Background.

The S.D.I.project was started in April 1968 with the construction of a thesaurus and basic design work. In January 1969 a trial service was started with 30 users. The same service became operational in July 1969 and is still growing. Input is cumulated monthly into a Review of Current Literature.

2. Coverage.

Subject:- Allthe subject interests of the G.K.N. group of companies and subsiduaries in the U.K. and overseas. This is mainly engineering components and metallurgy.

Forms:- 150 journals are selectively covered with occasional items selected from a further 350 publications. Patents and a few reports are also covered. Patents amount to around 60% of the total input.

Foreign material is abstracted and indexed in English.

3. Input.

Items are abstracted (indicatively, maximum of 500 characters), indexed (maximum 20 terms) and allocated a subject heading code. Indexing is controlled by a thesaurus of 1,800-2,000 terms which was based on the E.J.C. and A.S.M. thesauri. The data sheets are keypunched in the computing department.

Information:- Subject heading code, authors, title, bibliographic reference, abstract,and keywords.

Volume:- Input amounts to around 100 items per week.

4. Profile construction and match strategy.

Profiles are constructed by the Information officer from user statements as follows:-

Statement of users interests(narrative form).

Additional information(definitions, formulae,trade names etc)

Six document references directly relevant to interests.

Profiles are modified as necessary. User feedback is via preaddressed reply postcards. For each notified item the user marks up :- of interest, please send copy

of interest, no copy required

of no interest

comments.

Profiles consist of keyword terms linked by weighting, that is the simulated Boolean logic of Brandhurst(17). Each profile is restricted to ten fields or questions of ten terms.

5. Output.

Weekly. Output would appear to be about one month out of date for British journal material, about a week for British patents, two months for German patents and five months for Russian patents.

6. Users.

At present there are about 200 users in U.K. and overseas installations of the G.K.N. group of companies. Eventually about 1,500 users are hoped for.

7. Statistics

Information from user reply cards is collected. Data on the number of profiles in different installations and the number of hits per profile and per run is also collected.

8. Costing

None at present.

9. Computing.

An I.B.M. 360/30 is used with programmes specially written in PL 1 and Assembler languages.

Comments.

This service should become well established and expand to provide a vital service to its users. It is to be hoped that when costed it will still be felt worthwhile and continued. Perhaps part of the data base will be made,available outside G.K.N. This would be a good thing as it covers many subjects not covered by other data bases.

References.

FISHER M.E. Private communication.

International Business Machines (I.B.M.)

Hursley Park,

Winchester.

1. Background.

I.B.M. can probably claim to be the originators of S.D.I. Their S.D.I.system has gradually developed as their computers capabilities have increased. The Current Information Selection (C.I.S.) section of the I.B.M. Technical Information Retrieval Centre (I.T.I.R.C.) is decentralised with I.B.M. personnel in Eurppe being served by the La Gaude processing centre in France. The C.I.S. data base is prepared at I.T.I.R.C. at Yorktown Heights, N.Y. and sent to the European centre for dissemination.

2. Coverage.

The data base covers, I.B.M. documents, selected non I.B.M. documents, and I.B.M. invention disclosures.

3. Input.

Tapes received from I.T.I.R.C. are processed at La Gaude. Document details and abstracts are keypunched in upper and lower case variable word length format.The computer checks validity, sequence and spelling.

4. Profile construction and match strategy.

Profiles are by information personnel from data sheets provided by users. These give details of position held and a short paragraph describing the activities for which information is required.

Feedback from users is via Port-a-ppnch cards supplied for each reference. Users mark or punch out the appropriate box:- a) Abstract of interest/document not needed.

b) Send copy of document.

c) Abstract of interest have seen document before.

d) Abstract not relevant to profile.

Comments are invited.

Users return these cards to their local library who supply the documents and consider the need for profile modification. Profile modifications are forwarded to the European centre.

Profiles consist of words or word strings linked by Boolean logic or NOT and MUST logic. Technological terms and author names can also be used as search terms. Item files are matched against the whole profile file simultaneously.

5. Output.

Output is three times a month in the form of computer printout, with accompanying user response cards for each reference.

Full bibliographic references are supplied with an abstract and a list of word matches.

6. Users.

In 1965 there wase 1,200 users in the U.S.A. and250 served by tho European centre.

7. Statistics.

Evaluation of user responses revealed of 75-80% over the period February to September 1965.

This is based only on response cards returned. Only 59% were returned. It would be difficult to estimate whether cards were not returned because the references fell into catogory d), or c) or a). The relevance factor was derived using a),b)and c).

8 Computing.

The I.B.M. 360/ computer now used allows for remote terminal entry of data, remote enquiries, direct access storage devices and time sharing. This means an improved performance and swifter and easier profile modification.

Comments.

The high relevance figures are typical of such a user related data base. The I.B.M. programmes are available and suitable for other in house sytems.

References.

KAUFMAN S. System techniques and applications. August 1966 ITIRC-008.

LUHN H.P. A business intelligence system. in I.B.M. J. of Research and Development.

LUHN H.P. Selective dissemination of new science information with the aid electronic processing equiptment. in Amer. Doc. 1961 12 (2) 131-138.

MAGNINO J.J. C.I.S. A computerised normal text current or awareness technique. November 1965 ITIRC-006.

Shell Research Ltd.

Woodstock Agricultural Research Centre,

Sittingbourne,

Kent.

1. Background.

Since 1962 Shell at Woodstock have been developing a fully integrated mechanised information storage, retrieval and dissemination system. This was originally mainly concerned with internal reports and chemical structure processing. As part of the system published literature was scanned and relevant references subject coded with a five digit numeric code were processed to produce a current awareness bulletin. Index cards were also produced for users who wished to keep their own specific subject files. In order to expand the service without extra staff C.T. tapes were received. The tapes were received from January 1967 and the service on them started in April. S.D.I. output for individual profiles is collected and used to form subject reference files. The subject reference files are merged together to form a current awareness bulletin.

2. Coverage.

Profiles are run on a variety of subjects, not on chemical subjects. These include computer science, microbial chemistry, mass spectrometry, organic, chemistry, biochemistry, pesticides, and toxicology.

3. Input.

C.T. tapes with some biological material added from journal scanning. This added material accounted for 20-30% of the S.D.I. notifications. C.T. tapes were to be replaced by CBAC and C.A.C. tapes in 1969.

4. Profile construction and match strategy.

Profiles are constructed by information staff from disscussion with users. Users suggest their own profile modifications.

Profiles are of a similar form to those used by U.K.C.I.S.

5. Output.

Fortnightly, index cards for S.D.I. users. Other outputs from the system include a composite awareness list , subject reference files and feature cards. 5"x8" card output is planned.

6. Users.

All users are on site. There are 170-200 possible users of which 40 were running profiles at the end of August 1968.

9. Computing.

Originally C.A.S. programs were being used and processing was done on the I.B.M. 1401 computer at the headoffice in London, forty miles away. The system ran into trouble in August 1967 because the number of search termsof the profiles allowed by the software was being exceeded. This meant that the size of the profiles had to be reduced and new users meant further reductions. A Univac 9300 was to be installed in August 1968 at Woodstock and new programmes written for this were expected to solve the problems.

Comments.

This shows the feasibility of running an S.D.I. service as part of a larger information system for a fairly small clientele.

References.

DAMMERS H.F. Integrated information processing and the case for a national network. in Information Storage and Retrieval 1968 4 (2) 113-131.

DAMMERS H.F. Phased transition from non-mechanised information storage and retrieval to on-line computer operation. in Mechanised Information Storage and Retrieval. Proceedings of F.I.D./I.F.I.P. joint conference. Rome 1967, North Holland 1968.

GALLAGHER P.J. Some experiences with S.D.I. in Inf. Sci. 1968 2 (3) 103-106.

Standard Telecommunications Laboratories Ltd. (S.T.L.)

Automated Information Dissemination System (A.I.D.S.)

Harlow,

Essex.

Project leader - R.D. Kerr-Waller.

1. Background.

In 1965 S.T.L. decided to test the feasibility of an S.D.I. type service for their workers. This project was run by one of the research staff,not a member of the information unit or the library staff. An experimental system was set up which included the testing of alternative methods of indexing. After the initial experimental phase of two to three years the system was dropped and never became perational.

2. Coverage.

Subject:- Electronics.

Forms:- Abstracts from Science Abstracts section B for the nine months June 1964 - February 1965 were used together with research and development reports and patent abridgements.

Volume:- Around 10,000 abstracts, 150 RandD reports and 108 patents.

3. Input.

Literature was indexed by three different types of indexer who had a free choice of descriptors. Details of items were punched onto paper tape for input to the computer. The descriptors were compared with the cluster thesaurus containing 10,000 words and allocated the concept number of the appropriate cluster.

4. Profile construction and match strategy.

Profiles are constructed by the users using the subject thesaurus,containing 6,000 words, to give ideas 6or descriptors.

A one hour lecture was given to users on profile construction. Only three test profile modifications were carried out.

Descriptors in the profiles were arranged in three groups. Group one was used for the main keywords and the other two groups contained descriptors which qualified these in the main group. Profile descriptors were converted to concept numbers by the computer and then these were compared with the literature concept numbers for possible matches.

6. Users.

336 engineers at S.T.L.

9. Computing.

Programs were written for the experiment and run on ITT ADX 7300.

Reference.

KERR+WALLER R.D. Automated Information Dissemination System (A.I.D.S.) technical evaluation. in Inf. Sci. 1967 1 (2) 91-105.

U.K.A.E.A. Atomic Energy Research Establishment.

Harwell,

Didcot.

1. Background.

Since the end of 1968 the Information office of the Harwell library has been receiving Nuclear Science Abstracts (N.S.A.) on tape and has launched an S.D.I. experiment using these tapes. The aims of the experiment are "to demonstrate the feasibility of setting up certain services using the N.S.A. tapes, to obtain user reactions to pilot scale services and to carry out comparisons with other services"(ref. below). A manual S.D.I. service eas already in operation and information bulletins and specialised subject bulletins were also being produced by the library.

2. Coverage.

Subject:- Nuclear science within the fields of chemistry, earth sciences, engineering, instrumentation, life sciences, metals and ceramics, physics and reactor technology.

Forms:- The same material is covered by the tapes as is covered by the hardcopy though more detail is present on the tapes. N.S.A. covers journal articles,reports, books, patents, conference proceedings and theses.

Input to N.S.A. comes from atomic energy bodies in Canada, Denmark, Finland, Japan, Norway, Sweden, U.K. and U.S.A.

3. Input.

N.S.A. tapes are produced by the U.S. Atomic Energy Commision Division of Technical Information Extension (D.T.I.E.) as part of the computerised production of Nuclear Science Abstracts.

Tapes contain full bibliographic detail and keywords instead of the abstracts of the hardcopy form. Only the keywords are searchable.

At present the system cannot deal with an input of mose than 100 items per run. This has caused trouble and will probably be extended to a limit of 180 items.

4. Profile construction and match strategy.

Information staff construct profiles based on user statements as follows:-

General description of nature of work, subject, discipline, theoretical or applied. Present use made of N.S.A.hardcopy

Likely number of abstracts of interest in each N.S.A. issue

Detailed description of work including:-

Specific materials, methods and processes.

Possible synonyms in use in the literature for all specific terms.

For general terms which aspects are and which are not of interest.

Indications of whether design and performance ofequiptment and special enviroments are of interest.

The N,S,A, category numbers where topics might be expected to be found and those where they will be found but not wanted.

Abstract of interest from recent N.S.A. issues or titles of interest from other sources.

Profile modifications are based on user feedback. Users rate notifications according to the following categories:-

a) of immediate interest.

b) of general interest.

c) of no interest.

d) cannot decide interest on

the basis of information given.

To get some idea of recall, a sample of users check through a hardcopy of N.S.A.. This copy has to be circulated and provides rather a limited response but has been useful for analysing unexpected misses.

Profiles consist of questions made up of groups of selectors linked by Boolean logic. The selectors are chosen from the Euratom thesaurus of 16,000 words as the N.S.A. keywords are.

5. Output.

Fortnightly. N.S.A. tapes are received in advance of the hardcopy version. Previously only one hardcopy was obtained by airmail. Thus the dissemination method is speedier. Material takes a minimum of two to three weeks to appear on the N.S.A. tapes.

Output is on 6"x4" cards giving, N.S.A. iss. no., N.S.A. abstract no., title, author, research establishment of origin, bibliographic reference including inclusive pagination, date of issue, language of article and the keywords assigned to the item with matched keywords marked.

6. Users.

Members of A.E.R.E. staff thought suitable who after interview were still thought suitable. Some external users are to be included. In June 1969 there were 46 regular internal users and 16 external user profiles being prepared.

7. Statistics.

The following were based on 31 profiles run against 6 issues

The average number of documents retrieved per user was 24 (60 was the maximum number allowed by the system). The average relevance was 60% (71% of users scored an average relevance of at least 50%.

A questionnaire was sent to users who had received notifications from six issues up to the tenth issue of 1969. The results from these questionnaires were as follows.

All found the service at least of some use and were willing to continue participating. The keywords provided with the references were only found moderately useful and the marking of matched keywords was not especially appreciated. Most users were fairly satisfied with their notifications and if anything prefered fewer references. Most users felt the service had picked up documents they would otherwise have missed.

8. Costing.

Tapes are free as U.K.A.E.A. provides the U.K. input for N.S.A.

9. Computing.

An I.B.M. 360/65 is used with programs in Fortran and Assembler languages by A.E.R.E. staff. Teething troubles have been mainly concerned with space allocation foe runs.

Comments.

At present the "N.S.A./S.D.I. service provides more comprehensive coverage to,those users who previously depended on the information bulletin and special subject bulletins but is less accurate and selective than the manual information office service."(ref. below)

It would appear that this experimental service works fairly well. It could be very useful to many external users Itis to be hoped that it will become externalycommercially available.in due course.

Reference.

TERRY J.E. Private communication.

U.K.A.E.A. Atomic Weapons Research Establishment (A.W.R.E.)

Aldermaston,

Berkshire.

1. Background.

An experimental S.D.I.service was run on C.T. data base in 1966 and 1967 using C.A.S. package programs. These ran into the same type of problems as Shell did with theirs though Aldermaston had the trouble first and sorted out the problem first. In 1968 the S.D.I.service became operational using National Research Council of Canada's(N.R.C) programs for the IBM 360 series. Aldermaston is infact now using N.S.A. tapes also (since August 1969).O.S.T.I. is supporting this experiment on title searching of N.S.A:.tapes for comparison with the keyword searching at Harwell. Inspec tapes are being reformated for title searching with the N.R.C. programs.

2. Coverage.

C.T.,N.S.A., and Inspec S.D.I. experiment tapes.

3. Input.

Commercial tape input which amounts to around 2,000items per month.

4. Profile construction and match strategy.

Profiles are constructed by information staff in discussion with users.

Profiles are modified as required. User feedback is via six monthly questionnaire and relevance rating of each notification. One of the output copies is marked up with either "R" for relevant or "B" for borrow and returned to the library.

5. Output.

Fortnightly, two copies of computer printout.

6. Users.

80 at present.

8. Costing.

Machine time costs 30/- per user per run.

9. Computing.

An IBM 360/50 is being used with N.R.C. programs written in PL 1.

References.

CORBETT L. Chemical Titles - S.D.I. trial at A.W.R.E. in J. Doc. 1967 23 (2) 150-151.

CORBETT L. Using commercially available literature tapes for current awareness services. in Inf.Sci. 1968 2 (3) 83-102.

CORBETT L. Private Communication.

Unilever Research Laboratories.

Colworth House,
Sharnbrook,
Bedford.

1. Background.

The Colworth and Welwyn research laboratories cover Unilevers interests in Foods, Animal foodstuffs, and Biological safety testing. Too few staff were available to cover the current awareness side of the information function so new methods were sought. Profiles had been run fairly successfully by C.S.R.U. on C.T., CBAC, and the C.A.S. S.D.I. abstract experiment but C.T. and CBAC were not felt to be the ideal data bases. In the middle of 1968 a feasibilty study was made into the use of I.S.I. tapes for an'in house' computerised S.D.I. system. The library bulletin service was still maintained in its original form independent of the S.D.I. system.

2. Coverage.

The multidisciplinary/coverage of the A.S.C.A. service

3. Input.

Only I.S.I. source tapes were used.

4. Profile construction and match strategy.

Profiles are constructed by information scientists in disscusion with users.

Profile modification is encouraged, users initiate any changes. High recall profiles are aimed at in the beginning with precision increased with subsequent modifications.

Profiles consist of words,authors names and journal titles linked by Boolean logic and weighting.

5. Output.

Weekly. Unilever estimate that I.S.I. tapes material is 20-30 days out of date when it appears on the tapes, takes 7- 10 days to reach the S.D.I. processing centre, and 2-3 days to run , making the S.D.I. service about 30-43 days after the publication of the original articles. This means that S.D.I. notifications are received one to two weeks after the item appears on the library shelves. The library bulletin in fact provides a speedier service for U.K. material though not for foreign material.

6. Users.

About 350 internal users using around 20,000 search terms between them.

7. Statistics.

Recall and relevance ratios were not considered to be appropriate measures. Instead the numbersnof hits per week and the number of relevant hits per week are recorded. The nelevance figures are very different for different users the highest being 78% . Outputs are often very small suggesting that recall may not be very good. The data base is considered not to be so suitable for the Food technology users as the Animal research users as the average weekly numbers of hits for these are 8 and 24. However, the relevance measurements are 39% for the Food technologists and 17.7% forthe Animal research users.

8. Costing.

The I.S.I. source tapes cost $ 8,000 p.a. The costs of the service are £20,000 which works out at 1/6 per hit and 6/6 per usful output item.

9. Computing.

An average of 200 minutes computerptime per run is used.

The IBM 360/50 with 256K bytes core stoee at Port Sunlight is used via a remote access system based on IBM 2780 terminals. New and updated profiles are input from the terminals and the tape search is initiatde from the terminals. Because of the restriction on on-line transmission speeds it is uneconomic to print large quantities of output at the terminals so output is printed at Port Sunlight and delivered by courier to the research laboratories.

A great deal of thought went into the choice of software. At present National Research Council of Canada (N.R.C.) programs designed for C.T. data base are being used. A separate conversion program is used for converting the I.S.I. records into C.T. format which is written in Assembler language. Possibly in the future the IBM ITIRC/TEXPAK or the IBM Document Processing System (D.P.S.) will be used with modifications.

Comments.

One of the reasons for using the I.S.I. tapes was their complete coverage of journal contents. This meant that additional journals could be covered by the system either by in house keypunching of added input or by addition of another tape data base to fill in the gaps.

References.

ROWLANDS D.G. Private communication.

SHAW T.N., BUCK E.R. Mechanisation of current awareness services.at Unilever Port Sunlight..in Inf. Sci. 1967 1 (3) 91-108.

Derwent Publications Ltd.

Derwent Publications Ltd. have since 1963 been producing current wareness publications for different branches of the chemical industry, aimed to serve commercial interests rather than pure research interests. From January 1980 the following services will be available.

Ringdoc.

300 journals are covered selectively for the pharmaceutical industry. Informative abstracts are provided in the form of profile booklets in 42 profile subjects.

Pestdoc.

A similar service is provided covering 170 journals for manufacturers of pesticides and agricultural chemicals.

Vetdoc

Assimilar service covering vetinary literature.

These three services are available in several forms. which are:- abstract journals; abstract books, suited for photocopying; transparency books, suitable for dyeline or diazo reproduction; abstract booklets(thematic booklets) where abstracts are divided into small subjects (thematic groups) or microfilm reels.

Also available are a series of services covering patent literature as follows. These services cover patent specifications published in France, Japan, Switzerland, U.K., U.S.A., U.S.S.R., Western Germany, Belgium, Eire, Netherlands, South Africa, Australia, Canada, East Germany and Sweden which amounts to 83%of the worlds specifications. Several output forms are available:- company code cards(for searching on individual companies inventions); manual code cards allocated subject classes (270 chemical structure classes and 120 pharmaoological and agricultural products and activities punch code cards; and magnetic tapes(quarterly for retrospective searching). Programs are available for the I.B.M. 360/40. Custom searches are availble at £20 per question.

Farmdoc.

Patent specifications of pharmaceutical or vetinary interest

Agdoc.

Patents covering pesticides, herbicides, fertilisers and and other similar agricultural chemical topics,

Plasdoc.

Plastics and poymers patents.

Organdoc.

Organic chemicals patents.

Central Patents Index.

This service includes all the above services material together with other subjects. The total subject field covered is divided into six divisions each divided into a further twelve sections;-Plasdoc; Farmdoc; Agdoc; Food and detergents; Chemdoc;.Textiles and paper and cellulose; printingand photographic coating; petroleum and chemical engineering; nucleonics, explosives, protection; refractories, ceramics; metallurgy. These sections are further divided to form 122 classes which can be related to the International Patent Classification.

E.S.R.O./E.L.D.O. Standard Profiles. (S.P.)

These are standard subject S.D.I. bulletins issued monthly in a similar way to the ordinary S.D.I. service. There are 111 subjects or S.P.(69 for technology and 42 for science). The cost of the service depends on the number of S.P.s taken, for the first five topics it is £6 p.a. per topic.

E.S.R.O./E.L.D.O. Databank.

This is a new service commencing in January 1970 dealing with electronic component quality data and new electronic products of European manufacture.

New electronic product information is fed into the system and matched against user 'profiles'. Users are notified biweekly of new products in 35 categories or on specific components of interest to them. The users pay a subscription which means that no charge is made to manufacturers for inclusion. The service covers all member states of E.S.R.O. and E.L.D.O., that is Belgium, Denmark, France, Germany, Italy, the Netherlands, Spin, Sweden, Switzerland, and the U.K.

Component Quality Data is again disseminated selectivly to users. Users receive test and reliability reports on components eminating from nationally approved bodies such as the French C.C.T.U. and the Swedish F.T.L.

The layout of the new product information sheets is similar to that of the Indata service notifications. It is on sheets and includes , specification data, free text description of the item, manufacturers data and graphical data.

INDATA, Automation, Instrumentation Data Service.

1. Background.

Indata Ltd. a subsiduary of Guest Keen and Nettlefolds Ltd. in close cooperation with the Scietific Instruments Research Association set up the Automation Instrumentation Data Service in January 1967.

2. Coverage.

More than 2,000 manufacturers supplied input of their new products free of charge in the fields of electronics, automation and instrumentation.

3. Input.

Information from manufacturers was set out on 80 column card giving details of, name of product, manufacturers name and address, detailed engineering specification, free text description of up to 800 words with photographs and line drawings where necessary to the description of the product.

4. Profile construction.

Users could choose from 309 subject categories to describe their interests. The computer sorted out new product data relevant to users interests as described by the subject categories.

5. Output.

Users received current cards matching their interests as they were produced and a weekly accessions list covering their interests. The cards were kept in special binders in pockets forming a modular catalogue and service girls visited users to file these. Punched card reply stubbs formed part of these cards and these could be returned via Indata to the manufacturers who then furnished the user with further information on the product concerned.

Users.

In May 1968 15,000 categories were being subscribed to

A facility for small users to link together and use a Central Technical library as centre was available. This reduced costs as the library received the full set of product cards for all categories and individual users just received a profile accessions list.

7. Costing.

£30 to £375 depending on the number of categories received £5 for apersonal profile list of up to 20 categories and 5/- for each additional category. If the special library scheme was used then the central library was chargdd £375 and users were charged £15 for up to 50 categories and 5/- for each extra one.

9. Computing.

An IBM 360/40 was used. A suite of 29 programs were used which took 16 man-months to writein Assembler language. The system required a minimum core store of 65,356 characters.

Comments.

This system promised great things it outwardly appeared successful and was rapidly gathering support. However after just over a year of operation it folded up. The use of a computer in this system was not entirely necessary.

References.

BARLOW D.H. Computer based information services. in Data Processing 1968 10 (2) 102-110.

Iron and Steel Industry Profiles(I.S.I.P.)

Iron and Steel Institute.

The Iron and Steel industry for many years published an abstracts section in the Institute journal based on the Institute library accessions. In 1960 they started the Abstract and Book Title Index Card Service (A.B.T.I.C.S.) which supplied filing cards of journal article abstracts and book titles with U.D.C. numbers for incorporation into classified subject files.

I.S.I.P. is intended for individual users rather than libraries. There are 20 subject headings to choose from The service covers more than 1,000 journals selectively together with books and BISSITS translations, indicative abstracts are provided for journal articles. Output is weekly in continuous sheet form with average output amounting to 4-10 items for the different subjects. The service is swifter than the abstracts section of the Institute journal. Costs are from £9 to £21 depending on which category is taken. There are reductions for taking several subjects. The cheapest subject at the lowest rate is £2 per annum. A photocopy service is also available and BISSITS can be purchased

Patent Office.

Subject matter tabulation service:- Weekly lists of the numbers of U.K. patent specifications in chosen classes of the classification scheme. Cost is one shilling per page.

Selected patent specifications:- Users can place a standing order for all patnts in chosen classes of the classification scheme. The cost is only that of the specifications which is four shillings and six pence each. The specifications are distributed as they are published.

Scientific Documentation Centre (S.D.C.)

Holbeath House,

Dunfermiline.

1. Background.

In 1960 Dr. P.S. Davison a Scots chemist set up a centre to collect scientific data and study its best exploitation. In 1962 a limited company was formed of individuals and organisations who used the services of the centre. The Centre now has a collection of data and documents which are exploited by Dr. Davison and a number of part time women graduates.

2. Coverage.

Multidisciplinary over the field of science and technology though the best coverage is in the fields of spectral and analytical chemistry. A large collection of spectra is also held.

1,940 journals are scanned together with the U.S. Government Research and Development Reports, Current Research and Development in British Universities and C.A.T.S, Current U.S. and British University Thesis Titles, book titles from major publishers, Mass Spectrometry Bulletin, Chemical Abstracts and other secondary services.

3. Input.

Graduates scan the above publications and allocate them to the different subject categories and gies them index codes. The system deals with around 373,000 items per year.

4. Profile construction.

150,subject categories are available some of the larger ones are subdivided and spectral and analytical subjects can be taken only for certain classes of compounds if desired.

5. Output.

OUtput can be weekly or monthly on 5"x 3" or 80 column cards. The L.S.E. considers S.D.C. to be about eight weeks behind C.T. and A.S.C.A.

6. Statistics.

A case study was carried out by the Centre comparing sources of information for chemists on " Applications of computers to Mass Spectrometry" References were collected over the period 1965 to 1968. The percentage of the total references found by each service were as follows:-

Analytical Abstracts	3 %
Bibliography of Mass Spectrometry	11.0%
C.A. Pink	10.5%
C.A. subject index	24 %
C.T.	11.5%
Chemische Zent.	4.5%
Mass Spectrometry Data Centre(M.S.D.C.)	32 %
Referativnyy Zhurnal Khimik	5.5%
S.C.I.	21.5%
Science Abstracts	8.0%
S.D.C	39 %
Miscellaneous	5.5%

On this pet topic of the Centre, S.D.C. gave the best results though M.S.D.C. came near and was not running for the full period.

7. Costing.

The lowest priced subject category costs £15 , most subjects are in the £15- 40 range though the really broad topics cost up to £250.

Comments

In a few cases references are picked up by this service which are missed by the computer services, this is often because the complete articles are often scanned when subject categories are being allocated. However, unless the users interests fit neatly into one of the subject categories a great deal of unwanted material will be received.

References.

DAVISON P.S., MATTHEWS D.A.R. Letter to the editor. in Aslib Proc. 1969 21 (7) 280-284.

DAVISON P.S. Private communication.

U.K.C.I.S. S.C.A.N.

Narrow subject current awareness publications are being produced by U.K.C.I.S. based on the output from broad profiles run on the S.D.I. service. So far the Selective Current Awareness Notes (S.C.A.N.) have been produced for Steroids, Radiation and Photochemistry, and Group VIII . These cost £12 per annum and fifteen more titles are planned.

The Experimental Information Unit in Oxford are developing similar bulletins. They are producing a twice monthly bulletin on Electrolytic solutions by running profiles on A.S.C.A. and C.T.and having the output edited by two chemists working in this field. Other subject bulletins are planned for Enzyme Chemistry and Catalysis and other topics not planned by U.K.C.I.S.

CHAPTER 3.

ANALYSIS OF S.D.I. SYSTEMS.

In this chapter the basic structure of a mechanised S.D.I. system will be considered. Some conclusions will be reachdd as to the effectiveness of the different methods used in S.D.I. as represented both in the systems surveyed in the previous chapter and those to be found operating in the U.S.A. The component parts of an S.D.I. system, used to describe the systems surveyed in the previous chapter, are considered separately. The effect of variables in components of the systems on types of users is considered here whilst the interaction of the complete system types, formed from these component parts, with user types is considered in Chapter 4.

3-1 Background.

In 1958 Luhn first published his idea of a "Business Intelligence System" (63). This abstract idea was developed into a well defined system in a later article (64) and from here I.B.M.s internal S.D.I. system for reports developed. By 1965 it had become a fairly well known and sophisticated system. At this time the American Chemical Society had started using computers in the production of their Chemical Titles publication and the Eli Lilly Company and the Olin Mathieson Chemical Corporation began to experiment with the use of C.T. in machine readable form as the basis of S.D.I. systems(39)(73). Other large industrial organisations and academic institutions in the U.S.A. had started to set up mechanised information retrieval systems and S.D.I. systems based on Luhns original broad concept(19)(78)(79).These are described in several surveys (26)(51)(12). In this country, interest in S.D.I. began around 1965 when initial investigations and experiments began. Some systems developed more quickly than others. 'In house' systems, being smaller, had the advantage here. The Chemical Society set up a research unit at Nottingham University(54)(55)(56)(88) which soon got an S.D.I.system underway based on the American Chemical Society's machine readable data bases. The National Electronics Research Council took a little longerwith no machine readable data base available they had first to form one. Thus since 1965, in this country, systems have moved from experimental to either operational or obsolete(e.g. S.T.L. and Indata) and new systems are continually being set up.

However, as S.D.I. is still a relatively new subject, not enough systems have been tried and tested for any formulae to be derived with which new designers can work out the optimum for their environment.and set up an immediately operational system. Thus every system requires an initial experimental phase. Whether this is extensive, with most parameters being tried a variety of ways before the optimum is decided upon and the system set up, or whether a system is set up by an arbitrary choice becoming operational when it can be made to work reasonably efficiently, or whether already established systems are copied, the experimental phase will need to be carefully nurtured and ofcourse financed. O.S.T.I. grants have been vital in the setting up of most U.K. non industrial systems. Such grants are only available for experimental work and systems must be developed fairly rapidly to a self supporting state. Industrial systems will probably have to develop even more quickly and prove themselves either of greater value or less cost than previous equivalent systems.

Although there is no one perfect S.D.I. system for all situations the U.K.C.I.S. system would seem to be sufficiently well developed and effective to be applied to most machinereadable data bases which are not controlled by a thesaurus. For in house situations and those where a special document collection forms the basis of the system, design will have to be tailored to the environment. It is to be hoped that O.S.T.I. grants will continue to be available for setting up generally available systems. It would be worthwhile O.S.T.I. also giving smaller grants to 'in house' systems which would be willing to make their machine readable data bases or sections of them, generally available, as this would facilitate more specialised information entering the general information network in a convenient form.

3-2 Coverage

The coverage of a system is a very important factor. What has not gone into the system cannot be retrieved. For S.D.I. to be worthwhile to most users the coverage must be good enough to satisfy the majority of the users interests. The user should be able to virtually depend on his S.D.I. service as far as his core interests are concerned. Many workers in the S.D.I. field claim that the user must still do his own browsing as well. The user should need to browse through primary journals only for ideas and to fill in peripheral interests. For the expense of S.D.I. to be justified it should provide an almost perfect service in the fields it professes to cover. In addition to this the user must know how comprehensive the service is and how much it can be relied upon.

Subject coverage.

Here arises a problem in the central organisation of S.D.I.services. The problem is firstly whether to be exhaustive over a small subject area, or to give a fairly good coverage over a larger, possibly multidisciplinary subject area. Services have so far developed in various depths resulting in some highly specialised services with a limited clientele such as the Plasma Physics system and others with shallow coverage over a wide area such as S.D.C. and A.S.C.A. It would seem unsatisfactory if the user needs to take more than one service to satisfy his needs.

The ideal might be thought to be a series of separate services developed in single disciplines with in depth coverage. However as far as journal literature is concerned, if Bradford's law of scattering is to be believed(16), only one third of items existing within a specific subject field will be found in journals specifically devoted to that subject. This means a great deal of duplicated effort will occur if separate systems attempting in depth coverage are set up for different subject fields. According to Connor (26) this scatter is greater in the applied sciences than in the pure sciences, e.g. in chemistry, three quarters of references come from 10% of the journal titles, while in the biomedical field only half the references come from 10% of the journal titles. Thus the current awareness problem in the applied sciences is far greater than in the pure sciences. A greater need for S.D.I. type services might be expected in the applied sciences, and of course greater difficulty is involved in providing such a service, perhaps this is why there is at present a scarcity of custom S.D.I. services in the applied sciences. This problem is discussed further in chapter 4 . To avoid the duplication of effort, one central all embracing service is the answer or at least just a few broad subject based services. Figure 1. shows the overlap of the subjects covered by custom S.D.I. services.

Form coverage.

As far as the depth , that is the amount of material covered, is concerned this should be taken to its maximum as opposed to its optimum. I.S.I.'s claims that the 2,000 key journals which they cover account for over 90% of signif cant scientific reports. This is not really a deep enough coverage though in I.S.I.'s case a multitude of other literature is covered through citations. Journals may be the most obvious source of scientific information but an S.D.I. service restricted to journal coverage will not satisfy many users.

Journal articles form only part of the mass of source material and to some users this will be a small part. Many journals contain only information on pure research and then usually quite some time before it is completed. Report literature tends to be more up to date and often is more concerned with applied research and development. Much report literature is not readily available when originally published, and even "available" items may be highly elusive. S.D.I. services must bring in this form especially as it is poorly covered by more traditional current awareness tools. Report literature could be dealt with more easily by a small specific subject centre, especially one with associated active research interests, as type of organisation attracts report information to itself and often copies of the actual reports by exchange agreements. The other necessity is patent information. In a highly competitive industrial environment, patents will be of prime importance. Apart from their obvious commercial interests they can be of great use to the pure research worker. Conference proceedings are also important. The research worker is useally quite well aware of conferences in his own field, but those in peripheral fields may be of equal importance to him. Again the special subject centre will easily be able to collect informationfrom this source. Books are important though they rarely contain original material. These should be included in an S.D.I. service and are fairly easy to cover well. Trade informationis also very important to many research workers but is difficult to deal with successfully in an ordinary S.D.I. system. Indata or the E.S.R.O./E.L.D.O. Databank type service would seem to be the best for this form of information. Most custom S.D.I.services cover only journals as shown in Figure 2.

The more obscure the material source the more important it is that it gets into the system, as this enhances the value of the service to the user.

Foreign coverage.

This is also true for material of foreign origin and foreign language material. Some form of international cooperation and interlinking of S.D.I. systems elsewhere is a hope for the future. Inspec publications and full data base not only gain U.S.A. input from cooperation from the American Institute of Physics and the Institute of Electrical and Electronics Engineers but cooperation with German, French and U.S.S.R. sources is expected. Inspec tapes will be experimented with in Canada, Belgium and Sweden, Poland and Japan and presumably these cooperating countries will eventually provide information to the system. A similar situation exists for Nuclear Science Abstracts tapes, where several atomic energy organisations in different countries provide input for their countries in exchange for the tapes. A similar link exists between the European members of E.S.R.O./E.L.D.O. and N.A.S.A. for S.T.A.R. and I.A.A. tapes. Medlars also works to some extent on this type of exchange agreement.

This cooperation solves only half the probàem; the next stage is to overcome the language difficulty. This is often done by providing abstracts in English for input to the system. Although time consuming and costly, this is probably necessary to provide sufficient information to the research worker for him to decide the possible need for a complete translation of the original. The S.T.L. system(57) of allocating index codes was partly done to enable input of foreign material, cutting out the language problem. The Abbot et al. (1) comparison between C.A. services and A.S.C.A. revealed that C.A. coverage of foreign material was superior, especially for the more difficult Russian and Japanese. A.S.C.A. tends to wait for translations to appear before including items and this means a substantial time lag as far as Russian cover to cover translations are concerned.

Contents coverage.

Even if complete coverage of material in a subject area is not possible the important thing is that it is well defined. This is particularly important when it comes to selective or complete coverage of contents of journals claimed to be covered. Custom S.D.I. services should state exactly what material they cover and stick to it, so that the user knows exactly where he is and can supplement the service where necessary. In a specific subject oriented system, many of the journals covered are only covered selectively and this means that the user whose interests are only peripheral to the service subject will find it almost impossible to supplement it. A.S.C.A.'s complete contents coverage is much appreciated by users especially as it covers all those other snippets of information contained in journals such as editorials, correspondence and corrections. Ideally a centralised service would cover all literature completely, either by relying on title only or using unslanted abstracting or indexing, making the item accessible to all users likely to be interested in it.

The coverage of an S.D.I. system should be well defined in depth, breadth and comprehensiveness.to enable the user to be sure where he stands. The desired quality of coverage will be better and more easily obtained when custom S.D.I. services are batter coordinated.

3-3 Input.

The type of input used determines the flexibility of a system and thus is very important.

Input method

There appears to be a trend towards using available machine readable data bases as inputs to S.D.I. systems. Where these tapes are the byproducts of other machanised systems such as computer typset publications and computer sorted indexes, costs will obviously be reduced. However, the convenience and cheapness may well be obtained at the expense of the detail provided, which may well not be in the most suitable form and may necessitate complicated software. Other problems are inherent in the remote processing of magnetic tape information files: these problems are described by A.K. Kent(54). Available machine readable data bases are listed in Appendix III. A description of,available machine readable information files appears as an appendix to L.Corbett's article on'Using commercially available literature tapes for current awareness services'(29). A more detailed description of available data bases appears in the report on 'Mechanised Information seervices in the University Library'(89). Most of the S.D.I.services surveyed in this dissertation use available machine readable data bases.

For a mechanised system the choice of input method is between these commercially available tapes, keypunching details of the available document collection or a combination of the two. Keypunching solely for S.D.I. input is going to make the system very costly, but presumably tapes willbe stored for use in retrospective searching and may well also be used to produce a broader subject current awareness bulletin for more general distribution. Most library systems serving research workers produce a current awareness bulletin (see Appendix I). If this is part of the information system little extra work may be required to use the same details as part of an S.D.I. system. This was how the Culham system developed and the current awareness bulletin is soon to be distributed commercially. This form of input to a system depends on a first rate document collection. Keypunching may be found to be too time consuming and costly and may complicate the software unnecessarily as verification and error-checking procedures will have to be incorporated.

As more publications become computer typeset, relying on commercial tapes for input is going to be more feasible. However, this means that retrieved items will be restricted to those that are gathered in the secondary publication net, unless an efficient method of adding supplementary keypunched material is devised. Going back to the idea of the central organisation of S.D.I. services, here is a case for the centralised collection of commercially available tape data bases with additional material keypunched into the system by specialised subject centres.

Input information.

If a commercial tape data base is used, the detail input to the system is going to be that required for the original publication. This may be restricted to title, such as for a KWIC index publication like C.T., or an abstract or digest may be available for searching as in the case of CBAC and POST. Indexes to these publications will provide further search terms such as author names, subject descriptors and perhaps source organisation. The type of subject index involved will also be important. If the allocated keywords are natural language selections as in C.A.C. or with Lynch's articulated indexing (65), one set of problems will be involved, while thesaurus-controlled indexing used in Medlars, Inspec S.D.I. investigation and E.S.R.O./E.L.D.O. and S.T.L., will involve other problems.

If a system is to provide its own input then the following questions must be answered by the system designers:-

a) Do titles provide a suitable search base?

b) If not is this best extended by searching an abstract or perhaps full text or is it better to allocate keywords for search terms?

c) If keywords are used should this be a natural language selection or should indexing be controlled by a thesaurus?

a) The value of titles as a search medium varies with subject field. Abbot et al. (1) found that biological titles were more informative than chemical titles. In their consideration of which items from C.B.A.C. hits would have been retrieved by considering title omly, 50% of biological items would have been captured compared with 31% of chemical items. Swanson(87) found 86% ofentries in Index Medicus already had subject index words or thesaurus related words in their titles. In the applied science field where the journals tend to be more of a trade type than learned society public-ations, titles are more catchy and meaningless. Information concious users will be familiar with the type of title appearing in their field and can construct their profiles accordingly. Similarly, information-concious authors may give their papers meaningful titles; hence titles may be more reliable in the future.

b) Search of the abstract or full text may be advantageous, but it also uses a great deal of computer time as well as initial keypunching time. It can also lead to difficulties. Abbot et al.(1) found that in C.B.A.C. searches individual words from different paragraphes unrelated in the abstract were occasionally linked by the computer to form an invalid match with word combinations in the profile. This difficulty can be overcome by separating the paragraphs and perhaps sentances at the input stage.

Indexing of any kind is costly and time consuming and obviates the use of the computer as highley qualified and therefore costly personnel are still required to fulfill this function. A time lag will also be introduced which should be avoided in a current awareness service.

The way of overcoming these disadvantages is by automatic indexing or abstracting using the computer. A lot of work has been done in this field in the U.S.A., including that of Hillman at Lehigh University and in the SMART project at Harvard. Although automatic indexing has been shown to be quite possible even with some semantic control, the problem of keypunching whole documents remains,as the character recognition facility in computers is not very far advanced.

c) The constraints on the system imposed by a thesaurus seem hardly desirable. The Cranfield experiments,using aerospace reports, indicated that far from retrieval being improved with a thesaurus as aid, the contrary was infact the case and natural language keyword selections produced the better results. The type of retrospective retrieval employed in the Cranfield tests would not display the same problems as those involved in running profiles on an S.D.I. system, and this was a specialised data base, but similar effects might be expected. It has been evident from the survey of S.D.I. systems and from user interviews (Appendix II) that several problems are likely to arise with thesaurus controlled systems.

A users interests can only be adequately expressed in thesaurus allowed terms where the users interests lie within the bounds of the subjects for which the thesaurus was expressly designed. This shows up in the Medlars service where users whose interests were not purely medical,that is were not only concerned with case histories of patients, tended to have difficulties in profile construction. This type of problem is likely to arise in thesaurus-controlled systems except where a small well-defined disciplie is involved and a limited clientele with known interests are served as may be found in an 'in house system such as G.K.N. Even Inspec with its well defined discipline has run into trouble because of its varied clientele (see Appendix II). The other main disadvantage of a thesaurus controlled system is that it is more difficult for users to construct their own profiles as the structure of the thesaurus must be fully understood.

A great deal of work has been done by the U.K.C.I.S. unit at Nottingham on the questions discussed above but little has been published. From what has been seen during the preparation of this dissertation it would appear that thesaurus control is not desirable for S.D.I. systems. Indexing and abstracting introduce a time lag which cannot be justified unless it increases recall substantially. This time lag could be reduced if more use were made of author abstracts and subtitles. Many people doubt the validity of author abstracts but certainly they have the same slant as the item making them a good indication of author intent. Sage (78)(79) of the Ames Laboratory S.D.I. system favours the use of author abstracts for S.D.I. input.

The amount of detail input is important to the user. The tape services provide the basics but in very different forms. Some standardisation is needed. An adhered to standard on author citation would make all the difference both to S.D.I. systems designers using borrowed and modified software and for systems where several data bases are used. The Ames Laboratory system(78) (79) have had this sort of trouble when using N.S.A. and I.S.I. tapes together. A consistent practice of author citation would solve a lot of problems. Source organisation names, authors cited,subtitles and date of issue of item are all very useful terms for searching on, and the full pagination is very useful both for estimating the importance of the article (as are the number of citations) and when photocopies are required. A.S.C.A. has shown the value of citations. Little additional effort would be required to incorporate these into a keypunched input system. As much detail as possible should be input in such a way as to make all terms searchable: tagging is probably the most convenient method of doing this.

For tape data base input, some standardisation is needed. For keypunched input, as much detail as possible should be input in a searchable form. The natural language search techniques are being perfected by U.K.C.I.S. and many of the problems solved. With more aids to profile compilation available such as the KLIC (Key Letter inContext) of U.K.C.I.S. for use in choosing word fragments and more 'ideas' thesauri available in narrow subjects, which may be produced by the Experimental Information Unit at Oxford running the Liason Scientist Experiment(L.S.E.) (60), natural language searching S.D.I. will be an attractive method. Titles or expanded titles ,as they become more meaningful should provide a sufficiently suitable search media, perhaps with a broad subject category allocated for use in excluding unwanted subject fields.

Input volume.

It might be thought that the larger the input the better for S.D.I. systems in order to achieve maximum recall. However, Bivona and Goldblum (12) devised a measure which they called the'dissemination ratio.' This measures the ratio of items matched by the system to those fed into the system. If these unmatched items can be kept to a minimum the system is considered to be more effective. This is rather similar to the balancing of recall and relevance but at the system level instead of the user profile level.

The limit to the input volume is imposed by software and core storage, and if input is too large it will have to be batched and the profile tape run through for each batch. Alternatively,as is probably more usual the profiles are placed in the core and the input item file matched against the profile file instead. Which way round this is will depend on the relative sizes of the input files and profile files. If Bivona and Goldblum's dissemination ratio is important in the effectiveness of a service, this has yet to be investigated, then a small, narrow subject service tailored to users needs would be more effective than a large centralised service. The Plasma Physics service's method of producing a major word glossary from the profiles is likely to be useful in keeping the dissemination ratio at a fairly high value. Figure 1 is an attempt to indicate the relative sizes of the custom S.D.I services in this country.

3-4 Profile construction and match strategy.

Methods of profile construction and modification.

There are three basic approaches to profile construction. The most appropriate method will depend to some extent on the system being used. In the first of these the S.D.I. system operators compile profiles from either user interview or user statement. Other systems employ intermediaries between the system and user: persons qualified in the broad discipline of the user and practised in the art of profile compilation do the constructing in consultation with the user. Liason Scientists were used for this in the Students Chemical Information Project (S.C.I.P.) (35). The third method is for users to construct their own profiles with the aid of search manuals and seminars as do users of U.K.C.I.S.

The Experimental Information Unit in Oxford E.I.U. (60) feel that users should participate as much as possible in profile construction but need educating in this. Some form of link is also necessary between the user and the system, especially as S.D.I. systems operators become more adept at overcoming profile construction difficulties. In the L.S.E. the E.I.U. are testing the effectiveness of different degrees of user participation in profile construction and modification by giving user groups different amounts of assistance in their profile construction and maintenance. The ease with which users can construct their own profiles will depend to a large extent on the complexities of the system. Users will be more able to construct their own profiles for natural language systems as then the profiles will be based on their own subconcious knowledge of the terminology used in their field. Thesaurus based systems will not be so easily dealt with in this way as the whole structure of the thesaurus must be understood by the profile constructor. It is not surprising that systems such as Medlars, Inspec and E.S.R.O./E.L.D.O. place the responsibility for profile construction on the system operators. This will also be the case in many'in house'systems where the systems operators understand their users interests. However, it is interesting to note that the Culham system operators have found that profiles constructed by systems staff from external user statements have performed no worse than these constructed from first hand knowledge and interview for internal users.

Almost more important than initial profile construction is profile modification, as this is bound to be necessary in the early stages of a new profile and as users interests change. User attitudes to S.D.I. have been found to suffer because their interests have changed somewhat and they have been helpless to modify their profiles, or because profile changes have been slow. It is quite useless for an S.D.I. user to be receiving output from a profile which describes interests which are no longer held. Profile modification must be easily done and immediate. Systems which rely on user feedback in the form of relevance assessments will often have too slow an effect. The user must know how the system works and be able to indicate to the operators where changes need to be made in his profile. The system which requires users to give a relevance rating to each item he is notified of, will irritate the user in time as his profile appears not to be giving the desired results. The Inspec

S.D.I. investigation shows up particularly badly here. Users do not appear to understand at all how the system works. They do not appear to have received copies of their profiles or modifications and the only way they have of initiating modifications is by sending back their relevance ratings faithfully with often unfavourable results. The Ames laboratory system uses automatic profile modification as does the IBM sytem to some extent. In the Ames system 'significance'values are attached to profile terms, these being a measure of the probability of matching. User response cards are processed by the computer and profile modification is automatically carried out. Users are penalised for not returning their response cards by temporary cuts in their S.D.I. service. The IBM system uses a'user pattern' system. The patterns are built up from response cards and old patterns are slowly phased out making profiles follow the users interests fairly well.

At the experimental stage, systems operators are keen to get their profiles perfected and will be inclined to employ user feedback to improve their system. When the system becomes operational or runs on a commercial basis then the onus is on the user to get value for money and it is up to him to be aware of how useful his references are to him and to modify his profile accordingly. By this time the user s should have a good knowledge of the system and be able to make the most of it. In the S.C.I.P. the Liason Scientists became fairly proficient at dealing with profile construction and the users gradually began to understand the system. In the final questionnaire 89% felt themselves capable of constructing their own new profiles and 69% felt capable of instructing other research workers in profile construction. However, what of the user who is not one of the early users? Information scientists and librarians should be trained in the techniques of profile construction so that they can instruct users and aid them in their initial profile construction and early modifications. Apart from the Liason Scientists of the S.C.I.P., the E.I.U. in Oxford are using six more local liason scientists of whom three are subject specialists and three library or information personnel. Also O.S.T.I. is sponsoring six "Information Officer" posts in University and College libraries at Birmingham, Imperial College, Salford, Strathclyde, Sussex and Cardiff. Their duties include "the promotion of the effective use of information services both traditional and mechanised (particularly by liason with specific subject based mechanised services)." (advert in Nature 26th. July 1969 xxx) It is to be hoped that these will not restrict their services to University members but will provide advice to all possible S.D.I. users in their regions. Dannatt (32) thought that University librarians should have been used instead of Liason Scientists in the S.C.I.P. It is to be hoped that in the future Information scientists on University library staff will be available to fulfill this type of function . At present University library staff have no time for such duties , perhaps O.S.T.I. will sponsor more posts.

As S.D.I. services expand, it will be necessary for users to be responsible for their own profiles. Information officers should be able to provide the liason function to some extent. and aids to profile compilation will have to be developed.

These should take the form of search manuals, KLIC type indexes and ideas thesauri. Swanson(87) in a controlled study of 50 question searches on a collection of 100 documents in the field of nuclear physics, found the overall average relevance of retrieved documents to rise from 68% to 86% when an ideas thesaurus was used to compile search questions. The L.S.E. is working on these profile construction aids.

Profile form and match criterion.

The pros. and cons. of indexing with or without a thesaurus were discussed in an earlier section of this chapter. In this section the possible methods of linking search terms to form profiles compatible with these input forms will be considered. Profiles are usually divided up into subprofiles or questions which cover different aspects of the user's interests. Questions consist of parameters taken singley or linked together in various ways. The parameters consist of lists of related descriptors, authors names, journal titles, citations or whatever search term types the system allows. Ideally,the different parts of the input are tagged and search term types used in the profile are labelled with their term type so that they are easily matched against the correct part of the input item.

Connor (26) in her survey of U.S.A. S.D.I. systems found weighting to be the most popular match method where a threshold weight is assigned as the criterion for a match. This would not appear to be the case in the U.K., at present, where the AND, OR and AND NOT of Boolean logic is used in most systems and weighting tends to be used to refine a Boolean linked profile. Brandhorst(17) describes how weighting and Boolean logic are to some extent interchangeable. However, both methods have inadequacies. It is not possible to express either of the logic relationships (A.B)+(C.D) or (A+B).(C+D) by weighting. On the other hand it is not possible to exclude items containing term X by using not logic without excluding those which contain term X togetherwith term Y when all items containing term Y are wanted, unless an overridding MUST logic is used. Both NOT and MUST will bring in a lot of unwanted material. According to Bivona and Goldblum (12) systems using MUST/NOT/MAY logic are ultra-sensitive to the balance between recall and relevance. A high degree of relevance may be obtained with a high loss of recal in such systems. They also state that the use of Boolean logic allows a high degree of relevance with good control over recall and that weighting yields a high recall with little sacrifice to relevance. A controlled test on Boolean logic versus weighting versus nunber of terms matching versus percentage of terms matching was done by Sprague (85) and is described by East (37).in his survey of S.D.I. Sprague took 'Miss'(relevantdocuments not selected) and 'Trash' (irrelevent documents selected) as his criterion for evaluating these matching methods.and users chose their desired ratio of Miss and Trash. It was found that weighted term logic performed best and the relative performance of Boolean logic decreased as the importance of recall increased. Weights have a further advantage. Different threshold weights can be attached to different document types and when large outputs appear the highest weight matches can be selected for printout.

The IBM system uses Boolean logic and also some NOT and

MUST logic. The Medlars system uses a series of subsearches in its profiles aimed at different depths of search for the same concept. This facility would seem useful in the early stages for refining a profile. The Culham construction of profile in matrix form is quite unique and requires some thought in compilation. The system operators feel that the complexity of the profile is necessary in providing the tighter control necessary for natural language searching on title only. This matrix method was built up rather by accident than from deep thought on information retrieval theory but it seems to work successfully in this narrow subject field. The main criticism of this complex profile form is that it is rather awkward to change, hence profile modifications are not actively encouraged.

S.T.L. profiles are formed by dividing search terms into three hierarchical groups. This provides a speedy matching method but as these are rather uncontrolled collections of terms they would be expected to bring in rather a lot of noise. E.S.R.O./E.L.D.O. S.D.S. use grouping as an alternative to pure Boolean logic links. This procedure is used for relatively simple logical relationships where the parameters consist of long lists of related search terms usually from the same set or generic group. By grouping some economies can be made in computer matching time. This is really a form of weight simulation of Boolean logic.

Boolean logic links between search terms with the use of weighting for difficult profiles would seem to be the best method, possibly with MUST logic used to bring in works by named authors and NOT logic to exclude items from journals which are regularly scanned. Instead of MUST and NOT logic weighting can be used in the foam of very large positive or negative weights applied to such terms. On-line profile manipulation should be available so that estimates of output size can be considered before valuable printout time is wasted and profiles made more or less specific as necessary. This facility would also make profile modification easier and more immediate.

3-5 Output.

Output frequency and uptodateness.

To be a current awareness service the frequency of an S.D.I. service must be greater than monthly. Ideally it should be weekly, but it rather depends on how important a part the S.D.I. service plays in the users' current awareness systems. If the user relies entirely on the service, apart from scanning a few key primary journals, then a weekly service may be needed especially if the research project of the profile is very much influenced by other work in the field. A fortnightly service is probably sufficient and if the service is three or four weeks post primary source publication anyway, another week makes little difference. This is presumably the attitude of Medlars monthly service in providing six months out of date information. The amount of output will also influence the desirability of a certain frequency of service. Too large an output at one time is likely to be daunting, whereas two or three references received each week when a relevant one only appears every few weeks is rather a waste of postage and envelopes. Another influence will be the availability of documents. If documents are to hand there is some point in keeping well uptodate with a frequent service whereas if a wait, running into months, is likely then a weekly service has no worth.

Most U.K. systems provide a fortnightly service with weekly services for some 'in house' systems where the documents are readily available. With more out of date services such as S.D.C. or Medlars a less frequent service is provided.

Output form.

It seemed apparent at the start of this dissertation that the best output form was filing cards but users who were interviewed showed little interest in the output form they were receiving and little enthusiasm for card output. This may well hame been a reflection of the user's poor information organisation and that they kept no personal card file. The S.C.I.P. questionnaire revealed criticism of the paper output received and requests for cards from some users. Some of the users interviewed for this dissertation and described in Appendix II, felt card output to be desirable. Others got round this by doing scissor and paste jobs, while some cut up their computer printouts and used them as temporary intermediate records, while tracing the actual documents and later transfered the details to cards when the relevence of the actual document had been assessed. In the case of normal all upper case printout of the computer users, when accustomed, found it easy and quick to scan and it would seem to be preferable for users with high output and relatively low precision when cards provide a costly content for the wastepaper basket. Cards are slightly more costly to provide but do not constitute any great problem for software or hardware. Both forms should be available to users as alternatives. If computer printout is used it should be in double column format to provide compactness for ease of scanning.

Output information.

Output should include author, title and full bibliographic reference which should include full pagination and date of publication. The language of the item is also very important to the user. The name of the organisation with which the author is associated is very important to the user for initiating research contacts and as they like to send off to the authors for reprints rather than rely on borrowed copies or purchased photocopies. Resnick's study (72) indicates that titles are as effective as abstracts for relevance judgements. However, users like abstracts Inspec, users, who were provided with the keywords of the item but apparently not informed what these words were on their output cards, seemed very keen to have abstracts provided. Index keywords do not appear to be of much use to users. If abstracts are input to the system they should be printed out at least for foreign language items so that users do not need to search out an abstract in one of the published journals.

Other outputs for the user.

Some systems provide their users with detailed statistics of their profile performances. These are particularly useful for remote users. A measure of the number of matches each profile term has over a period provides a useful indication to the user of his profile effectiveness. Bivona and Goldblum (12) called this the selection- participation ratio and considered this to be a measure useful for evaluating S.D.I. systems. Copies of user profiles should be sent to him after each modification and at regular intervals to make him aware of his profile and as a check that modifications are as requested. Most services at present involve the user to a small extent only as most of them are still experimental. By providing the user with such statistics he should be enabled to make better use of the service.

3-6 Users.

The majority of users of S.D.I. custom services in this country are chemists of whom most are in academic institutions This is because information techniques in the chemical field are better than in others and C.S.R.U. was the first service to obtain substantial government support. The majority of C.S.R.U.'s users have been in universities because S.C.I.P. enabled final year postgraduate research students on Science Research Council grants to experience S.D.I. It is not intended to discuss the S.C.I.P. in detail here but its effect on the development of and interest in S.D.I. in the chemical field has been considerable and is worthy of note.

The Students Chemical Information Project (35)(84)

In the 1967/68 academic year,500 final year Ph. D. students in chemistry and pharmacy departments, who were on S.R.C. grants, were provided with a free S.D.I. service from C.S.R.U. run on the C.T. and CBAC data bases. Six Liason Scientists, who were postdoctoral chemists and had been given some crash training in S.D.I. and associated information techniques, were used to liase between C.S.R.U. and the students. These Liason Scientists were responsible for educating the users, usually by an initial lecture or seminar and then helping them to construct their profiles.

The Liason Scientists were responsible for dealing with profile construction,modification and user feedback.

> "The main objectives of this project were educational; it was hoped that a large group of academic users would gain experience in the use of this type of computer-based service, and that they might learn more about the existence and use of conventional chemical and related information services. Secondary objectives were to gain operational experience in routine provision of computer-based services on a large scale, to investigate the feasibility of using specially trained research chemists to act as intermediaries between the users and the service suppliers, to learn something about the information habits of students, and (ultimately) to have users with some experience of computer-based services working in a large number of Government, academic and industrial ~~research~~=laboratories."

This project achieved its main objective in that many chemists are now quite experienced users of S.D.I, knowing what they want out of the service and just how much they can expect from it. The secondary objectives also seem to have been achieved to a large extent. Interesting results were obtained from the attempt to learn about the information habits of students. and these should be useful when considering the whole future of S.D.I. and its place in information provision. The last objective of 'spreading the word' will be a rather more long term project as many research students remain in universities as postdoctoral fellows for a time.

The S.C.I.P. questionnaire attempting to discover the students future occupations revealed that large numbers will be going abroad for this postdoctoral experience so some may sample U.S.A. S.D.I. systems' services.

Liason Scientist Experiment (60) (61)

As to the "experience in the routine provision of computer-based services on a large scale"and the feasibility of the use of Liason Scientists, these have been carried over to the follow up project where they have provided a basis of experience. The follow up project is the Liason Scientist Experiment(L.S.E.). O.S.T.I. have given a grant to the specially formed Experimental information Unit (E.I.U.) in Oxford for it. The objectives of the experiment are :-

"(i) To provide university research workers with practical experience in the use of computer-based and other S.D.I. services;

(ii) to investigate more effective methods of user education;

(iii) to evaluate the comparative performance of selected operational current awareness services;

(iv) to developeexperimental services in selected subject ares in response to user demand;

(v) to define procedures for searching natural language data bases and to develop profile construction aids (e.g. a thesaurus)."

This has meant that approximately 5,000 research workers in chemistry,pharmacy,biochemistry,pharmacology,microbiology, botany and biology etc. will be receiving S.D.I. services from one or more of C.T. CBAC and POST provided by U.K.C.I.S.

Medlars, A.S.C.A., S.D.C, Approximately 800 profiles are being run, many of these are for groups of research workers. The average number of users per profile is 4.2. As a result of this, most university chemistry research workers will have either first or secondhand knowledge of S.D.I. services. Most research workers are just receiving one of these services but a selected sample (approximately 70) are receiving output from C.T., A.S.C.A. and S.D.C, and are involved in the Comparative Evaluation Experiment involving these services and their own scanning of Chemical Abstracts hardcopy. Participants in the evaluation experiment receive all three services and are now at the stage where the profiles have been refined to give comparative notifications, which will continue for about six months. Not only will these users be providing valuable information for the evaluation experiment but will be having an opportunity of forming their own opinions of S.D.I. and of the individual services so that at the end of the experiment they will be experienced in the use of S.D.I. and know exactly what its worth is to them personally.

Users of other systems.

The Inspec S.D.I. investigation is giving experience of S.D.I. to research workers in all environments in the field of electronics. However, the user education provided by this service seems minimal and users do not seem to be learning much about the practicalities of S.D.I. They play no part in the construction of profiles apart from supplying statements of their interests and relevant references and providing relevance rating feedback. Thus they are not really in a position to judge the true usefulness to themselves of the service or S.D.I. in general.

The Culham laboratoey provides a service somewhat similar to that provided by a Specialised Information Centre and its users are a rather select section of the scientific community, half of which are workers at Culham itself. The external users include many ex-Culham workers. User convenience would seem to be the only advantage in the S.D.I. service over the bulletin emanating from the same data base.

Paying users of A.S.C.A. are not very abundant as the service is quite expensive and this rather cuts out the academic research worker. Several chemical industrial organisations subscribe to A.S.C.A. as does the National Physical Laboratory running twenty four profiles at present. A few engineering an electronics industrial organisations are also running A.S.C.A. profiles. One year free trials have been made available to a few research workers at different times.

Personal profiles versus project profiles.

Hoshovsky and Downie in their survey of operational and experimental S.D.I. systems(51) posed the question as to whether profiles should be designed to describe users interests or projects. Users' interests are likely to be, and should be diverse, with interests not restricted to their immediate research projects. These interests are likely to be difficult to define in single question profiles and the user may well find it difficult to pinpoint his interests. If the profile is of the project it will be fairly well defined and will be shared probably by more than one research worker thus appeaRing to reduce costs. The trend is towards the development of group profiles, possibly even linking similar groups together and serving them with narrow subject current awareness bulletins compiled from broad S.D.I. profile outputs. In this way the user will have to supplement the S.D.I. service by journal scanning to satisfy his peripheral interests and add the 'ideas' part which cannot be effectively provided by an S.D.I. service anyway. It would seem that project profiles provide the most effective use of S.D.I. and the one most likely to be supported by financers.

Who S.D.I. users should be and what they can expect is discussed further in Chapter 4.

3-7 Statistics and methods of evaluation.

The main statistics collected for an information system tend to be relevance and recall ratios, despite the meaninglessness of such measures shown by the Cranfield tests.

These theoretical measures divide the total collection of documents in the system into four groups:-

a) documents retrieved which are relevant.

b) documents retrieved which are not relevant.

c) documentsnot retrieved which are relevant.

d) documents not retrieved which are not relevant.

then the recall ratio is $a/(a+c)$.

and therelevance ratio is $a/(a+b)$.

Neither of these measures have much meaning when describing an S.D.I. system especially as the definition of relevence is non standard.

To measure recall for a search question the number of items in the system relevant to the search question must be known as well as the relevant items retrieved in the search.

The only way to do this for S.D.I. is to get the user to scannthe whole content of the system. This is just about possible for systems with hardcopy versions of the data base such as C.T. or CBAC or for systems which also produce a bulletin covering the whole data base. However, users can hardly be expected to do this scanning thoroughly for long and it is likely to be far from full proof. The alternative is to get some estimate of recall by asking users to cite papers relevant to their interests which they have heard about by other means and have not been retrieved by the S.D.I. service. This method only really gives an indication of where the S.D.I. retrieval method breaks down. Recall tests on systems in their early stages are useful for plugging the gaps in system performance but are not meaningful as a standard measurement for comparisons.

To measure relevance, users must mark up those items in their notifications which are relevant. This seems simple enough but the definitiom of relevance is tricky. At present, relevance is described by different systems in different ways, making a relevance ratio quite meaningless as a comparative measure. When ratings amount to a simple choice between relevant and irrelevant, relevance ratios might be thought to have some meaning but these judgements based on different output details will add another variable if these details provided by the output are different to those used in the machine matching. This point is discussed in connection with the comparative evaluation of current awareness services (61). As it is, many systems use relevance rating scales of two, three , four, and five choices, making these making these measures from different systems incompatible and incomparable.

Relevance and recall measures are not independent quantities. If relevance is increased, recall generally decreases. For a fairly good system and user combination the recall and precision ratios are linked as shown in Figure 3 , by the continuous curve. This is a curve of operation and by varying the profile specificity almost any desired point on this curve can be chosen for the operational point. Other systems and user combinations and subjects will have different operational curves such as the dotted curves in Figure 3 . The point of operation on the curve will be determined by the user's desires. The ideal is to be able to operate at the point desired. Systems can be designed so that they operate at different recalls and relevancies but the ideal point of Figure 3 is apparently unobtainable and the optimum curve for a given system will vary from subject to subject.

From this it can be seen that citing a particular average recall and relevance for a system is rather meaningless particularly as it is the user's personal choice whether to go for everything there is on the subject with possible enormous outputs and a lot of 'trash' or for a small highly relevant output which will have a poor recall.

It will be interesting to see how the measures employed in the Comparative Evaluation experiment which is part of the L.S.E., work. These measures are basically recall and relevance measures though with fairly tight controls. Measures for individual users for each service will be taken at different levels of specificity, detailed analyses of recall failures will be made and systems will be compared only by individual users' experiences with each.

Probably the best measure of a systems success comes from the users'response. In the long run this will be the basis for judgement as systems go commercial. Like all information systems S.D.I. services should be designed for the user to give him what he wants and must be flexible enough to satisfy a variety of user demands.

3-8 Costing of S.D.I.

According to Bivona and Goldblum (12) costs can be divided into two types, fixed and variable. Fixed costs consist of personnel supervision, abstracting and indexing personnel, clerical personnel, keypunching and paper tape operators, materials and contingencies. Variable costs consist of computer processing time, materials, reproduction etc. and are related to the number of users of a system. This is for a system providing its own input and would be slightly different for a system using machine readable data base input. However, it means a cheaper service can be provided where there are a large number of users. In the case of'in house' systems the personnel are probably available from the days when a manual S.D.I. system or current awareness bulletin was provided. The added costs of a mechanised S.D.I. system will be tape subscriptions, software and computing time. If the input is from keypunched document details, these would have been typed in the old days so little added effort will be required to keypunch them instead. For tis extra cost more users will be able to receive a better service from the same staff.

Costing of S.D.I. is a relative quantity, for an 'in house' system, S.D.I. costs must be related to the costs of previous systems for providing staff with a current awareness service both in terms of library personnel time and users' reading time. For a custom S.D.I. service from the user point of view, the subscription must be offset against journal scanning time saved and whether this was spare time or not and the importance of references supplied which would not normally have been found. From the custom service supplier point of view, costs will be related mainly to the output size as printout time is the lengthiest and therefore most expensive quantity. This may be done either by charging directly on output size as U.K.C.I.S. does, or by charging higher prices for the use of search terms which occur frequently in the data base and are likely to produce high output, as A.S.C.A. does. Some charging is also likely to, be done on the profile size and form as this will determine the search time used which is quite costly. However, for an'in house service which is available to external users some arbitrary cost figure may be fixed which will easily cover the added costs of external usage of the system. Commercial costs are likely to be fairly high certainly not less than £30 per profile which is rather high for an individual worker though perhaps feasible if one profile will serve a group of workers.

By linking groups of workers with similar interests and producing a narrow subject current awareness bulletin based on a fairly broad profile serving several groups of workers, costs can be brought down. These bulletins will still contain references with a high relevance ratio for workers and are reasonably priced. U.K.C.I.S. charge £12 per annum for their SCAN publications and E.S.R.O./E.L.D.O. £6 for their Standard Profiles. S.D.C.'s cheapest subject categories are £15 per annum. Both the Inspec S.D.I. investigation and the Plasma Physics Current Awareness service will be run on a cost recovery basis in 1970 and it will be interesting to see what charging system these use.

S.D.I. is expensive but if it is effective enough it will be thought worth the cost. Barkla in his assessment of S.D.I. services (8) produces some interesting figures for the cheapest methods of getting the maximum number of relevant references by a combination of custom S.D.I. service profiles and scanning of primary journals.

3-9 Computing for S.D.I.

Very little information was obtained on computer processing and output times for S.D.I. Bivona and Goldblum give details of computing times for some U.S.A. systems,in their survey. Bloemeke and Treu (14) found that C.T. searches could not be run on a time sharing mode as the time used was too variable and there were often faults in the tape records. U.K.C.I.S. have the whole day of computer time for their use alone when they do their runs and so do not have this problem.

In most cases existing hardware will be used by services and so far most U.K.systems seem to be using K.D.F. 9 computers. Programs may be bought from the manufacturers or packages obtained with the data base or from other existing systems. A lot of programing effort is required to write programs from scratch. Appendix IV gives details of some software packages available for S.D.I.

CHAPTER 4.

S.D.I. IN PERSPECTIVE.

In this chapter the place of S.D.I. in the information network of the U.K. is discussed. The current awareness habits of different types of research workers are considered together with the advantages to them of different types of S.D.I. service. From these basic considerations conclusions are reached as to the proper development of S.D.I. services in this country. Finally by consideration of government policy as indicated by O.S.T.I. and the Dainton report the probable development of S.D.I. services is discussed.

4-1 Whether research workers want or need S.D.I.

The demands and needs of research workers for information are rarely the same thing. It is a great pity that this is so. It is to be hoped that in the near future the majority of research workers will be sufficiently well educated in information techniques to enable each individual to know his own special needs, for it is only the individual that can know. Projects like S.C.I.P. and Inspec's user analyses can find out the general trends in information habits. However , individuals still need to approach the problem in different ways as their subjects, project environments and the information sources available will make their problems unique.

The information explosion is with us and there is likely to be worse to come. It is obvious that a research worker cannot go it alone and find all his own information and do the practical research as well. Before starting a project these days a research worker must do a literature search or at least make use of some fellow worker's literature collection or pick a supervisor's brain. From there on it is a case of keeping up to date. Many estimates have been made of the time needed for a research worker to keep abreast of current information. Some estimates have even stated that it takes more than a week's workimg time to keep up with that week's literature. All estimates indicate that the worker cannot keep up with the literature adequately and research as well, without some aids. The traditional methods are the use of secondary publications in a narrow field or at least subject divided so that the user can restrict his scanning to the relevant sections. Abstract journals do a good job particularly if there is one which covers the users particular topic. These will bring to his attention items from many sources with a summarised version to cut down his reading time. They will also bring in material which he would not normally come across himself. However, he will still have to read a lot of unwanted material in case it is of interest and bhe material will be somewhat out of date as the time lag of an abstracting journal is usually at least three months and more for foreign publications.

The more helpful publications for alerting the scientist quickly to new material are the current awareness journals. The scanning of these however, will be very time consuming and it will be difficult not to miss things. It will probably still be necessary to scan a few core journals. A research worker cannot rely on finding everything important to him in a few core journals as so much scatter exists. Abstract journals and especially current awareness journals do not exist in all fields though important primary journals often include an abstract section or literature review. So our research worker, if he remembers and if he has the time that week, will need to spend probably one afternoon a week in the library scanning primary and secondary publications and on top of this he still has to find time for reading anything of interest which he finds.

An S.D.I. service will put before the usee notifications of items in his field with a pretty high ratio of relevant items. It will take the user but a matter of minutes to pick out what is of immediate interest to him instead of spending tedious hours poring over indexes and contents pages. This will leave the user more time to read the actual items and to browse through journals to satisfy his peripheral interests. However, S.D.I. will also help users who have a poor literature collection available to them and will mean that items of known interest can be requested from elsewhere.

The academic research worker.

The academic research worker will have to do practically all his information work himself as very few university libraries step from the library function into the information function. The journals he needs may or may not be available in his university or college library. If his department is small or of recent origin, or his interests are in a new field for that university, or if he is in a technical college or polytechnic, the journal provision in his field will probably be rather sparse. Abstract journal provision in a well established university will probably be good, though current awareness publications may be few and the library may well not be keen to subscribe to a variety of the latter as they are of such short term value. Current awareness publications will not be of much use in a remote library.

In many research fields, publications as mentioned above play little part in keeping research workers abreast of current developments. If senior staff in the project group are well established in the field they meet their colleagues from other establishments, here and abroad, at meetings and conferences and will be well aware of what is going on in their field and will probably be on author's mailing lists for reprints. More junior workers may well get their information from above. Many academic research workers feel that they can keep fairly well up to date by this method alone and tend to scorn information workers and S.D.I. as being unnecessary. However, there is a place for custom S.D.I. for groups where whole sections of departments are working on the same broad projects; and the isolated research worker, especially at the more junior level, needs the help that an S.D.I. service would give. Perhaps the Science Research Council could finance custom S.D.I. profiles for its sponsored students.

Research workers in Research Associations.

Research association workers are usually predominantly experienced researchers au fait with the narrow subject field they are in. They will probably be a fairly small group, hence communications between staff will be good. A strong information department will be present which is an integral part of the organisation and the library will be furnished with most of the literature in that narrow field. Thus the individual research worker may feel no real need of S.D.I. from an external source, though broad profiles to cover the whole of the research associations interests may well be felt worthwhile by the information department as another source of input for their information system.

Research workers in Government establishments.

Government establishments should provide a case somewhat similar to research associations where research interests are well defined, well established and fairly long term. However, if they are large organisations, as they often are, there is a tendency for the library to be a rather remote organ, rather like the university situation. However these days the research worker and information officer roles have a tendency to become one and the same as research workers get interested in information and move their desks from laboratory to library. This is a very good tendency. In these large Government establishments departmental libraries tend to spring up. These are often sparsely equipt, often badly staffed, and lure the research worker away from the better central library which is often some distance from his laboratory. Here is a good case for group custom S.D.I. profiles for each project team. This would also suit small establishments with inadequate libraries. However, for a large organisation it may be more profitable for the central library to produce its own S.D.I. service for staff as is beginning to happen in the larger government establishments.

Research workers in industry.

In industry there is a completely different situation. The research worker will often have a very poorly stocked library, unless his organisation is large and has a large research department with associated supporting facilities a library and information officers. In the former case the research worker will have a hard time keeping abreast of current developments and may well do unnecessary experiments.

A custom S.D.I. profile would be useful here though it might be found too expensive. When the research department is a separate entity a good information officer or librarian will probably be available. The infoemation officer will be faced with the task of keeping the research staff up to date often from only a moderate literature collection. He will probably be running some sort of current awareness bulletin or manual S.D.I. service for the staff. Broad S.D.I. profiles would provide a back-up to the literature collection for the information staff. In a really large organisation with strong research interests, group custom S.D.I. profiles may be the answer or a central 'in house' produced S.D.I. service may be justified.

Research workers do need some form of S.D.I. service, but whether this is an 'in house' manual S.D.I. service (formal or informal), an 'in house' mechanised service, or profiles run on a custom S.D.I. service, is rather irrelevant to the research worker. Which form is best for him will depend on the organisational framework of which he is part and on his subject field.

4-2 The form of S.D.I. service research workers should receive.

An S.D.I. service may be available but if it only provides a small portion of the users' needs it will not justify its costs. This partial appropriateness is particularly evident in the custom S.D.I.services and until these are more varied, flexible and effective, few research workers will find a use for them.

An 'in house' service is obviously specifically designed to fit its users needs. It is probably not possible for a nationally available custom service to be equally appropriate. In a close knit organisation such as an industrial organisation, research association or Government establishment an 'in house' system would seem to be best. If the organisation is small an 'in house' manual S.D.I. system is probably best, though if interests are well defined custom profiles may prove better. This 'in house' manual service could be boosted by a broad custom S.D.I. profile or one of the quasi S.D.I. services such as the SCAN service. For a looseknit organisation such as universities and some large government establishments a series of individual or group profiles run on a custom S.D.I. service are likely to be best.

U.K.C.I.S. now handling the sales of C.A.S. tapes in this country and in charge the C.A.S. custom S.D.I. service run by the former C.S.R.U., recommends purchase of the tapes for input to'in house' mechanised S.D.I. systems only for organisations who would need twenty five or more profiles of the custom service to cover their interests.

So that 'in house' mechanised systems can be set up with ease where applicable, suitable software must be available and perhaps more tape data bases. So that custom S.D.I services can be used where appropriate, these must be more numerous, varied and flexible.

4-3 Developments needed in custom S.D.I. services.

Custom S.D.I. services at present are unfortunately few and far between. Interest in S.D.I. in this country is rapidly growing. As can be seen from Appendix I, industrial research departments find a place for it and the trend would seem to be towards a growing use. As one academic participant in the evaluation experiment of the L.S.E. explained, if S.D.I. is not vital now, then it will be very shortly as the volume of literature produced increases.

As indicated in the previous sections of this chapter the custom profile form of S.D.I. service is necessary for several types of users but only users in a few subject fields can get the required service. Figure 1 shows how the custom S.D.I. services cover the field of science and technology. Pure chemists can receive a fairly successful and reliable service from U.K.C.I.S. on C.T. and specialists on the organic and biological fringes have POST and CBAC services. Chemical and biological workers with more general and diversified interests have A.S.C.A. However, industrial u users will have to supplement these custom S.D.I. services with some form of patent current awareness service unless POST-P is applicable to them. Derwent publications may provide the answer here. Workers in the pharmacy and medical fields are not quite so lucky as CBAC is not very good for pharmacists and Medlars is so out of date that it is relegated to a backup service. For pure biologists CBAC is good in some fields though it seems more difficult to handle than C.T. Biological Abstracts Previews may fill the gap between CBAC and the medical field, with Food Science and Technology Abstracts covering the applied biology field, assuming that custom S.D.I. services become available based on these. It is to be hoped that a custom service will be made available based on Nuclear Science Abstracts as this would help to fill the gap between separate services in physicsnand chemistry. The Inspec S.D.I. investigation has been covering electronics and particularly the semi-conductor side. It is to be hoped that the full Inspec tapes will become the data base for a custom S.D.I. service these tapes will cover the whole range of physics, electronics and electrical and control engineering. It is to be hoped that the E.S.R.O./E.L.D.O. Databank service will develop well to provide a commercial component S.D.I. service It promises to take the place of the short lived Indata.

This leaves the whole field of mathematics, mechanical engineering, production engineering, metallurgy and civil engineering with no custom S.D.I. service. Chemical engineering is fairly well covered by the C.A.S. services and the Engineering Index - Plastics tapes, at present being experimented with in several European countries, may lead to a custom S.D.I. service. If G.K.N. release their data base or even part of it as commercially available tapes, this may help in the forgotten engineering fields. E.S.R.O./E.L.D.O. ofcourse cover a wide variety of subjects. The situation holds promise for a fairly good coverage of most disciplines though a lot of subjects will be left in the gaps between. Only with natural language searching methods would systems seem to be flexible enough to cater for the likely varied clientele.

4-4 The organisation of custom S.D.I. services.

At present the government influence on S.D.I. organisation is centred around U.K.C.I.S. at Nottingham University. This is the organisation with all the know-how on custom S.D.I., having established procedures forC.T. and CBAC and then POST tape searches, these are being applied to other data bases. Tests are going on with Chemical Abstracts Condensates.

Biological Abstracts Previews and soon Food Science and Technology Abstracts and assistance is being given to Inspec in designing a custom service on the full Inspec tapes. It rather looks as if the trend is going to be towards developing a centralised services. Perhaps U.K.C.I.S. will concentrate on the chemical/biological side and Inspec will develop services on the physics/electronics/electrical engineering side. It is to be hoped that Engineering Index tapes will be included in some custom S.D.I. service. A custom S.D.I. service on N.S.A. would be desirable. This would probably be bst produced by U.K.A.E.A. Medlars should be left for what it was designed for: that is a retrosoective search service which it does well. The Plasma Physics service is of limited value since it has such a narrow subject base. It should be regarded mainly as an'in house' service providing for a few external users. External users especially in this country will be almost as well served by the by product, the Plasma Physics Bulletin,soon to be commercially available.

By initial sponsoring of S.D.I. projects, O.S.T.I. has some control over them and may well be able to develop the services into a unified system in this country. In the U.S.A., S.D.I. services are mainly centred on large organisations either industria or academic. European countries are only just beginning to take an interest in S.D.I., as already mentioned Inspec and Engineering Index tapes are being experimented with in various places. The use of C.T. tapes in Denmark is described by Boman (15). It is interesting to note that even in the U.S.S.R.'s unified information system (68),agriculture, architecture and civil engineering, medical subjects,standards and measurements, and patents all fall outside the main organisation VINITI, which provides information services in science and technology. So perhaps an S.D.I. service embracing all disciplines is not really the best. Also in the U.S.S.R. information system, there is an upward flow of information from individual laboratories and factories in the form of mini reports which are chanelled via the Central Industrial and Regional Information Centres for inclusion in the main abstracting journals.

Perhaps we should consider something on the U.S.S.R. lines in the U.K. for adding information to the present tape data bases of custom S.D.I. systems. For a good S.D.I. service, and S.D.I. has got to,be good to be of any use, extra material needs to be added to the basic machine readable data bases before input to custom S.D.I. systems. As already mentioned more foreign material, patents and reports are needed. This extra material could be fed in from Specialised Information Centres, Research Associatios, and other special subject centres who attract information on their specific topics of the less conventional forms. Some method of incorporating world wide patent information into S.D.I. inputs is needed as only then will custom S.D.I. be of much use in industrial enviroments. G.K.N. have a patent input amounting to 60% of their total input. Perhaps the patent office could take some part in the input to centralised custom S.D.I. services, otherwise agreements with Derwent publications might be possible. This would form a two-tier input to custom S.D.I. systems.

The other way of incorporating this extra information is by a two-tier dissemination system. In this case a central agency would disseminate information in fairly broad subject fields to specialised information centres of various types including the present S.I.C.s and R.A.s, who after keypunching in any additional material would disseminate information selectively to individuals and small groups of users, on their narrow profile interests. This could work where there were enough special subject centres in a unified network. But this we do not have at present so perhaps the system considered in the previous paragraph is more applicable to the situation in this country. This would also overcome the user's difficulty in deciding which special subject centre is best suited to provide his needs. Specialised Information Centres seem already to be moving t towards the two-tier dissemination system as they use broad profiles od custom S.D.I. services to provide part of the input for their own systems.

The Dainton report (Cmnd 4028 1969.) commebts related to S.D.I. and associated services seem to be restricted to paragraphs 413-415 and recommendations 63-66. A National Bibliographic service (N.B.S.) is recommended which would coordinate the country's bibliographic effort and take over the work of O.S.T.I. in this field. The N.B.S. would set standards for machine readable information systems. However, decentralisation is recommended in appreciation of the need for subject specialists in this type of work. Thus it would seem that the implementation of these recommendations would lead to the setting up of more S.I.C.s using the two-tier dissemination method described above. The hope expressed by the report was for the creation of a "netwokk of compact computerised services all directly accessible from user's terminals throughout the country."

This would seem to be an ideal situation but is unlikely to be the situation in the immediate future . Remote terminal document display is also a fair way off, in the meantime a first rate document supply service is required as a backup for the S.D.I. systems developing. At present the N.L.L. provides a satisfactoy service for the type of material notified by custom S.D.I. services. If more obscure material references are notified to S.D.I. users as it is to be hoped they will, then some additional document supply facilities will be necessary. If S.I.C.s are to add additional specialised material to the standard data base they will have copies themselves and should be prepared to provide a photocopy service to users for this specialised material or deposit copies at the N.L.L. for them to provide the service.

This is an exciting time for S.D.I. and a crucial one as well, for if organised properly at this early stage, this country could provide itself with a well coordinated information network. This could consist of a national network of specialised information centres covering the whole of science and technology fed by machine readable input from the computer produced publications of scientific and technical articles and by reports and patents and other items from originating bodies and the patent office. This input could come not only from this country but enter the system by exchange agreements, from other countries. This collection of documents and references would be disseminated in various ways, in small amounts to individuals via custom S.D.I. profiles, to groups in broader profiles or in SCAN type bulletins. This mass of data could also be available in large portios to organisations to form the bases of their own internal S.D.I. systems. These files of information would be collected and merged to provide bases for retrospective retrieval. Figure 4 attempts to represent this national information network.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

The outstanding feature of S.D.I. is that it is expensive. Added to this is the fact that S.D.I. service performance is often not nearly as good as it ought to be. The result is that user reactions are often not very favourable.

It is thought that some form of S.D.I. is necessary in many of the situations discussed in Chapter 4 and that it is desirable in many more. S.D.I. is going to have to come and the sooner the better. The problem is how are users going to be able to find the subscription and how can costs be reduced. On the first count users and ultimately administrators, who hold the purse strings, must be made aware of the true value of S.D.I. Many research workers are already experiencing such services and in time 'the word' good or bad will be spread. S.D.I. costs will be less when the right type of S.D.I. service is used in the right situation, again education and experience is needed, this time of information officers and librarians. Also computer costs should decrease and procedures will be developed which will cut down on computer time. Thus with larger numbers of users of each service the cost for individual users should be less in the future.

The following conclusions were reached on S.D.I. service operation. As the setting up of an S.D.I. service needs to be done with a great deal of careful planning it is thought that grant aid to new custom S.D.I. services is essential while they are experimental and is the best way to get them established. It is also thought that some grant aid , probably on a smaller scale should be given to 'in house' S.D.I. systems which are willing to make their internally produced data bases or parts of them, generally available. This would help to improve the variety of the data bases available. Custom S.D.I. services should cover more non-periodical and foreign material and their coverage must be well defined. Ideally tape data base formats should be standardised to make multidata base services feasible and allow for greater ease of addition of supplementary material. Software will probably have to be designed to overcome this non-standardisation.

It is thought that only with natural language searching methods such as those employed by U.K.C.I.S. and others, can a custom S.D.I. service hope to be as flexible as is necessary. However, not enough work has been done on thesaurus based systems to prove this fully. Ultimately, the user is thought to be the best person to compile his profile and initiate the necessary modifications. To make this method really work well more profile construction aids are needed in the form of construction manuals, and ideas thesauri and advice must be readily available to the user from systems liason officers or local information officers. These aids are gradually appearing and information workers are becoming more informed on S.D.I. techniques and more able to provide the advice required. Profiles seem to be able to express users interests most accurately when they consist of natural language words, word fragments and phrases linked by Boolean logic and weighting, with author and organisation names and journal titles as possible additions. To get profiles spot on, which is very important, requires modifications in the early stages and then occasional further changes to adjust the profiles to users ever changing interests. This is most easily done where on-line facilities are available and these are therefore highly a desirable part of a S.D.I. system's hardware.

An S.D.I. system is first and foremost a current awareness service, therefore it must be up to date and frequent. The form of the output should be flexible so that users can receive the form which suits their needs best. As much detail as possible should be given in the item notification to help in the tracing of documents. However, it is thought that abstracts are only necessary for foreign language material or particularly obscure material where more indication of content is likely to be worth while, before much time and effort is expended on obtaining the complete item. The measure of a S.D.I. system's success would seem to come from the user reactions which judge the ability of a system to be flexible enough to satisfy a variety of user's needs.

S.D.I. services are virtually useless without a good document supply to back them up. This document supply is dependent on the availability of documents nationally and internationally and the availability to the individual user of a library service which will obtain these documents for him. Unless the user can obtain most, if not all, the relevant documents brought to his attention by the S.D.I. service, in a relatively short space of time, then the S.D.I. service is wasted on him.

Assuming this document supply to be available to the user, then his need for S.D.I. and for the different types of S.D.I. service will depend on his work, subject field and approach and his environment. Custom S.D.I. services are considered to be particularly worth while for individuals in large loose-knit organisations or small isolated ones and for newcomers to a field such as Ph.D. students. It is thought that S.R.C. grant aided students should receive a portion of their grant to pay for a custom S.D.I. profile. In many ways, 'in house' S.D.I. services, mechanised or manual depending on the circumstances, are thought likely to serve the users needs more successfully than custom services, though probably at a greater 'real' cost. Broad profiles on custom S.D.I. services are thought to be the answer for boosting small 'in house' systems while tape data bases form the core of the input for the larger 'in house' systems.

Until the whole library and information field in this country and throughout the world is more co-ordinated efficient and effective, S.D.I. services will not be comprehensive and hence cannot really fulfill their function completely: that of providing users with notification of everything that is relevant to their interests, that they ought to know about. An S.D.I. service that falls far short of this ideal will not be a great deal of use and will justify its costs to few users.

Thus a S.D.I. service is thought to be necessary to many people in this country but as available at present it is only really worthwhile to a few and is rather a luxury to most. Unless custom S.D.I. services improve in their flexibility and comprehensiveness within a field then they will remain a luxury to the few who can afford them and the few who get them free. Before a great deal of effort and money is expended on S.D.I. services, the basic document supply methods in this country should be improved.

BIBLIOGRAPHY

1. ABBOT M.T.J., HUNTER P.S., SIMKINS M.A. Current awareness searches on C.T., C.B.A.C. and A.S.C.A. in Aslib Proc. 1968 20 (2) 129-142.

2. AITCHISON T.M. et al. Developments towards a computer based information service in Physics Electrotechnology and Control. in Information Storage and Retrieval 4 (2) 177-186

3. AITCHISON T.M. Investigating S.D.I. in Aslib Proc. 1965 17 (12) 343-350.

4. ANTHONY L.J., CARPENTER D.H., CHENEY A.G. S.D.I. using KDF 9 computer. in Aslib Proc. 1968 20 (1) 40-64.

5. ANTHONY L.J. CHENEY A.G. WHELAN E.K. Some experiments in the selective dissemination of information in the field of Plasma Physics. in Information Storage and Retrieval 1968 4 (2) 187-200.

6. ASPINALL W.K. Selective dissemination of information using a large computer. in Program 1968 2 (4) 135-148.

7. BARHYDT G.C. The effectiveness of non-user relevance assesments. in J. Doc. 23 (2) 146-149 1967.

8. Assessment of information services. Letter to the editor. in Aslib Proc. 1969 21 (9).

9. BARKLA J.K. Construction of weighted term profiles by measuring frequency and specificity in relevant items. Presented at 2nd Information Conference on Mechanised Information Storage and Retrieval Systems. September 1969. Cranfield.

10. BARKLA J.K. The University of Sheffield Biomedical Information Project. in Inf. Sci. 1969 3 (1) 13.

11. BARLOW D.H. Computer based information service in Data Processsng 1968 10 (2) 102-110.

12. BIVONA W.A., GOLDBLUM E.J. S.D.I. Review of selected systems and design for Army technical libraries. August 1966 AD 636 916.

13. BLEE M. INSPEC Project. in Data Systems April 1968 38-46.

14. BLOEMEKE M.J., TREU S. Searching Chemical Titles in Pittsburgh time-sharing system. in J. Chem. Doc. 1969 9 (3) 155-157.

15. BOMAN M. The Scandinavian user's attitude to information retrieval on computers. - What is wanted of new services. Experiences from the use of an S.D.I. (C.T.) system. in Levels of interaction between man and information A.D.I. 1967 306-310.

16 BRADFORD S.C. Documentation. (2nd ed.) London 1953 Crosby Lockwood.

17. BRANDHORST W.T. Simulation of Boolean logic constraints through the use of term weights. in Amer. Doc. 1966 17 145-146. (correspondence)1967 18 49.

18. BRANNON et al. Automatic Literature Alerting Service in Amer. Doc. 1969 20 (1) 16-20.

19. BROMBERG ,DUBINSKI ,PETERSON Bonneville Power Administration S.D.I. program. in Special Libraries October 1967 569-575.

20. CAWKELL A.E. Search strategies using the Science Citation Index. in Computer based information systems by B. Houghton 1968 Clive Bingley.

21. Chemical Society Reseach Unit in Information Dissemination and Retrieval. Search Manual.

22. CLAGUE P. The S.D.I. investigation in Electronics. in Computer based information retrieval ststems . ed. B.Houghton 1968 Clive Bingley.

23. CEVERDON C. The Cranfield Tests on index language devices. in Aslib Proc. 1967 19 (6) 173-193.

24. CLEVERDON C. Final report on the first stage of an investigation into the comparative efficiency of indexing systems. Cranfield Aslib Research Project 1960

25. COLUMBO D.S., RUSH J.E. Use of word fragments in computer based retrieval systems. in J. Chem. Doc. 1969 9 (1) 47-50.

26. CONNOR J.H. Selective dissemination of information a review of the literature and issues . in Library Quarterly 1967 37 (4) 373-391.

27. COOPER M. Current information ideas and practices. in J. Chem. Doc. 1968 8 (4) 209.

28. CORBETT L. Chemical Titles- S.D.I. trial at A.W.R.E. in J. D0c. 1967 23 (2) 150-151.

29. CORBETT L. Using comercially available liturature tapes for current awareness services. in Inf. Sci. 1968 2 (3) 83-102.

30. DAMMERS H.F. Integrated information processing and the case for a national network. in Information Storage and Retrieval 1968 4 (2) 113-131.

31. DAMMERS H.F. Phased transition from non-mechanised information storage and retrieval to on-line computer operation. in Mechanised Information Storage and Retrieval and Dissemination. Proceedings of FID/IFIP joint conference Rome 1967. North Holland 1968.

32. DANNATT R.J. Letter. in Nature 1967 216 208.

33. DAVIS C.H. S.D.I.- A program for Public libraries in Amer. Doc. 1967 18 (3) 139-145.

34. DAVISON P.S.,MATTHEWS D.A.R. Letter to the Editor in Aslib Proc. 1969 21 (7) 288-284.

35. Department of Education and Science. Office of Scientific and Technical Information. Students Chemical Information Project. Final report Parts 1and2 Feb. 1969 and June 1969.

36. EAST H. Design of a computer-aided information service for Physics with minimum concessions to the machinery. in Inf. Sci. 1967 1 (3) 107-119.

37. EAST H. Development of S.D.I. services. in Aslib Proc. 1968 20 (11).

38. EGGINS H.O.W.,WILLSHER M.J.D. Economic problems in running a Specialised Information Centre in Inf. Sci. 1969 3 (1) 3.

39. FREEMAN R.R. et al. Automatc preparation of selected title lists for current awareness services and as annual summaries. in J. Chem. Doc. 1964 4 (2) 107-112.

40. GALLAGHER J. Experiences leading to the introduction of an S.D.I. system.in J.Chem.Doc. 1966 6 (3) 151153.

41. GALLAGHER P.J. Some experience with S.D.I. in Inf. Sci. 1968 2 (3) 103-106.

42. GARFIELD E.,SHER I.H. A.S.C.A.-A new personalised current awareneess service for scientists. in The American Behavioral Scientist 1967 X (5) 29-32.

43. GARFIELD E.,SHER I.H. I.S.I.'s experience with A.S.C.A. in J. Chem. Doc. 1967 7 (3) 147-153.

44. GOFFMAN W. On relevance as a measure. in Information Storage and REtrieval 1964 2 201-203.

45. HARLEY A.J. MEDLARS. A comparison with hand searching in Index Medicus. in Inf. Sci. 1968 2 (2) 59-72.

46. HARLEY A.J. MEDLARS in context. in J.Royal College of Physicians of London 1967 1 (3) 261270.

47. HARLEY A.J. The Medical Literature Analysis and Retrieval System - MEDLARS. in Computer Based Information Retrieval Systems.ed. B.Houghton 1968 Clive Bingley.

49. HARLEY A.J. U.K. MEDLARS Information Rettieval Service Handbook for Users. N.L.L.

50. HINDSON The dissemination of published information to the executives of a major steel company. in Aslib Proc. 1965 8-22.

51. HOSHOVSKY A.G.,DOWNIE C.S. Selective dissemination of information in practice. Survey of operational and experimental S.D.I.systems.September 1967 19p. AD 668 072

52. JAGER H.D.M. The Mass Spectrometry Bulletin and Data centre. in Inf. Sci. July 1968 25-32.

53. KAUFMAN S. Systems techniques and applications. August 1966 IBM Technical Information Retrieval Centre ITIRC-008.

54. KENT A.K. Problems of remote processing of magnetic tape information files. in Aslib Proc. 1968 20 (11)502-508.

55. KENT A.K. Retrieving chemical information by computer. in Chemistry and Industry 7th. September 1968 (36)1214.

56. KENT A.K. United Kingdom experiences inthe operation ofaa retrieval and dissemination service based on Chemical Abstracts Service search tapes. Paper presented before the Chemical Literature Division of American Chemical Society September1968

57. KERR-WALLER R.D. Automatic Information Dissemination System (AIDS) technical evaluation in Inf. Sci. 1967 1 (2) 91-105.

58. KOLDER,HANSJOEG,SIMKINS S.D.I. and academic science libraries. in College and Research Libraries.1967 28 53-57

59. LANCASTER F.W. Evaluation of the MEDLARS demand search service. National Library of Medecine Jan.1968 278p.PB 178 660

60. LEGGATE P. Liason Scientist Experiment 1968/1970 Paper presented at O.E.C.D.Paris conference March 1969.

61. LEGGATE P. Proceeduresfor a comparative evaluation of current awareness services. Experimental Information UNIt Oxford. 3rd.July 1969.

62. LINE M. Large scale information retrieval systems available in the U.K. Bath University of Technology Library.

63. LUHN H.P. A Business Intelligence System. IBM J.Research and Development. October 1958 2 314-319.

64. LUHN H.P. Selective dissemination of new scientific information with the aid of electronic processing equiptment. in Amer. Doc. 1961 12 (2) 131-138.

65. LYNCH M.F. Subject indexes and automatic document retrieval. in J. DOc. 1966 22 (3) 167-185.

66. MAGNINO J.J. C.I.S.- A computerised normal text curret awareness technique. November 1965 ITIRC-006.

67. MARTIN M.D., SMITH J.R. INSPEC a new concept in information services. in Electronics and Power Feb. 1969 15 66-69.

68. MIKHAYLOV A.I. et al. Organisation of scientific and technical information in the communist world. 1966 AD627 802

69. MOUNTBATTEN,EARL.Controlling the information explosion in J. Institute of Radio Engineers April 1966 31 (4) 195-208.

70. National Electronics Research Council. N.E.R.C.Prpposal to investigate S.D.I. in Nerc Review 1965 1 (1) 3-9.

71. Office for Scientific Information policy III -Specialised Information Centres. in OSTI Newsletter 1967 (1) 2-3.

72. RESNICK A. Relative effectiveness of document titles and abstracts for determining relevance of documents. in Science October 6th. 1961 134 1004-1006.

73. RICE C.N. Acomputer based alerting system for C.T. J. Chem.Doc. 1965 5 (3) 163-165.

74, ROGERS F.B. MEDLARS operating experience at University of Colerado. in Bulletin Medical Library Association 1966 54 (1) 1-10.

75. ROLLING L.N. A computer-aided information service for science and technology. in J. Doc. June 1966 22 93-115.

76. ROSSMASSLER S.A. Modification of dissemination channels for scientific information. in J.Chem.D0c.1969 ~~0~~ 9 (1) 17-19.

77. ROWLANDS D.G. Unilever Research S.D.I. in Information Storage and Retrieval December 1969.

78. SAGE C.R. et al. Adaptive information dissemination. in Amer. Doc. 1965 16(3) 185-200.

79. SAGE C.R. Comprehensive dissemination of current literature. in Amer. D0c. 1966 17 (4) 155-177.

80. SAVAGE TR. The interpretation of S.D.I. data. in Amer. Doc. 1967 18 (4) 242- 246.

81. SHAW T.N.,BUCK E.R. Mechanisation of current awareness services at Unilever Port Sunlight in Inf. Sci 1 (3) 91-106.

82. SOMERFIELD C.A. Computer based information services in Aslib Proc. 1968 20 (12) 542-550.

83. SOMERFIELD C.A. State of the art of computer based information retrieval systems. in Computer Based Information Retrieval Systems. ed. B. Hougton 1968 Clive Bingley.

84. SOMERFIELD C.A. Students Chemical Information Project in Chemistry in Britain 1968 4 (2) 71-73.

85. SPRAGUE R.H. A Comparison of systems for S.D.I. Bureau of Business Research, Graduate School of Business Report no. 38 Bloomington Indiana 1965.

86 SWANSON D.R. Science journals and information of the future. in American Psychologist Nov. 1966 21 1005-1010.

87. SWANSON D.R. Searching natural language text by computer. in Science 132 1099-1104.

88. VEAL D.C. An experiment in the selective dissemination of chemical information. in Computer Based Information Retrieval Systems. ec. B. Hougton 1968 Civo Bingley.

89. University of California. Mechanised Information Services inthe University Library. California University Los Angeles Inst. of Library Research. Dec. 1967 PB 178 441

90. WAGNER A selective current awareness system using Engineering Index Plastics data base. in J. Chem. Doc. 1969 9 (2) 80-88.

91. WENTE V.A., YOUNG G.A. Selective information announcement systems for a large community of users. in J. Chem Doc. 1967 7 (3) 142-147.

APPENDIX I

Current Awareness Provision for Research Workers.

An attempt was made to get an overall picture of how research workers are kept abreast of current information in their field and in particular how many manual S.D.I. systems are in existence in research establishments, what use is being made of available S.D.I. systems and to discover any mechanised systems that had not been found by other methods.

a) Forty-eight research departments of industrial organisations were sent the questionnaire shown at the end of this appendix.

b) Thirteen of the larger Research Associations were contacted in an attempt to discover their information methods both for internal and external members and their use of custom S.D.I. services.

c) The main Specialised Information Centres were examined for their output methods and for their experiences of custom S.D.I. profiles for providing input.

a) Industrial Organisations.

The forty-eight industrial research departments were selected from 'Industrial Research in Britain' (6th ed. 1968 Harrap Research Publications) on the basis of the number of qualified personnel employed. Research departments of more than one hundred such personnel were included together with others which it was thought ought to be information conscious, such as chemical organisations and some which from library/information department job advertisements appeared to have an interest in S.D.I. These latter two categories were used in order to discover use of S.D.I. services and to discover any unknown 'in house' S.D.I. systems.

48 questionnaires were sent out. 41 replies were received of which 38 questionnaires were fully answered.

From these it was discovered that G.K.N. were running a mechanised S.D.I. system and this was followed up. Also 3 other research departments claimed to be planning mechanised systems.

15 library/information units were running 'in house' manual S.D.I. services for staff by scanning incoming documents. 11 of these were for more than just a few users. Another 2 were about to start running manual S.D.I. systems.

3 departments subscribe to Derwent publications and selectively distribute the abstracts.

33 library/information units circulate journals to staff.

36 library/information units produce a bulletin of some kind. All but 5 of these include individual journal articles, many were subject divided.

10 research departments were running profiles on one or other of the services surveyed in Chapter 2-1. This amounted to a total of 67 profiles being run.

Thus 27 out of 38 research departments had some form of S.D.I. system running or will shortly have and in addition to this S.T.L. had one previously.

b) <u>Research Associations.</u>

10 research associations supplied details of their information systems.

<u>Input</u>:- 2 were or were about to run profiles on A.S.C.A. (British Ceramic and British Welding)

3 claimed to be running a manual S.D.I. system for their research workers. Many of the others were too small to be running a formal system.(British Ceramic, British Iron and Steel, British Food Manufacturing Industries)

5 provided some sort of current awareness service from the library. These included bulletins of abstracts, xeroxed contents pages, current literature reviews, library accessions lists.

<u>Output</u>:- All obviously disseminated reports to members. Some produced more general abstract bulletins and current literature reviews and news bulletins.

3 had some form of selective distribution to members. The Rubber and Plastics R.A. (RAPRA) has a fairly sophisticated current awareness system.

4 claimed to run a document supply service.

The R.A.'s from whom information was received were:- British Ceramic; British Coal Utilisation; British Coke; Cotton, Silk and Man-made Fibres; British Iron and Steel; Wool Industry; Rubber and Plastics; British Scientific Instruments; British Welding; British Food Manufacturing Industries.

c) <u>Specialised Information Centres.</u> (SIC)

Biodeterioration Centre, University of Aston.

Biomedical Information Project, University of Sheffield.

Information Centre on High temperature processes, University of Leeds.

Mass Spectrometry Data Centre, AWRE Aldermaston.

<u>The Biodeterioration Centre, University of Aston.</u> (38)

Director - H.O.W. Eggins.

This is one of the early S.I.C., starting in 1965 and is fairly well established though its staff is still small. It now covers a broader subject field than it did in the beginning. Output from the centre consists of a variety of publications. The centre provides a collecting and storage unit for literature in its field. Input to the system comes mainly from a network of co-operating specialists scattered throughout the world. The Centre experimented with CT and Medlars S.D.I. profiles as input to the system but found them to be of little use. ASCA was also used but a sufficiently broad profile was too costly and so the service is no longer received.

<u>Biomedical Information Project, University of Sheffield.</u>

(8), (9), (10). Director - K. Barkla.

The Biomedical Information Project is concerned with collecting material for its monthly bulletin. This is done by scanning of core journals, scanning of Index Medicus and running S.D.I. profiles on CT, CBAC, ASCA and Medlars. A lot of work has been done in conjunction with the

Postgraduate Library School at the University of Sheffield on profile compilation and the relative effectiveness of the different input sources. The centre is also now co-operating with UKCIS and is running a profile on Biological Abstracts Preview Tapes. The S.D.I. profiles on CT, CBAC and ASCA were each found to add 5-10% to the collection of references.

Information Centre on High temperature processes, University of Leeds. Project Leader:- Dr.L.Barker.

The High temperature information bulletin is produced by the centre six times a year. Input for this is obtained from Plasma Physics bulletin and French & U.S.A. publications in the field. Current contents (Physical Sciences) is scanned together with U.S.A. and British R.& D. report publications, as well as a number of core journals.

Mass Spectrometry Data Centre (52). AWRE Aldermaston.

This Centre was set up at the end of 1965 and its output is via a monthly bulletin, price nine guineas p.a. Input to the system is obtained from various sources including the scanning of 180 journals and S.D.I. profiles run on the AWRE service that is on C.T. and N.S.A. data bases, A.S.C.A. and the C.A.C. experiment of UKCIS.

References.

Department of Education and Science O.S.T.I. Five Specialised Information Centres. Report by Aslib. OSTI report no. 5050. Oct. 1969.

APPENDIX II

User Survey.

Case studies of Custom S.D.I. users.

Various users of the different S.D.I. services surveyed, were interviewed concerning their experiences with S.D.I. An attempt was made to interview a few users of each service. These users tended to be in academic research, unfortunately, as these were easier to contact and discover in the first place. However, some industrial users were discovered.

The interviews were on the whole guided by the following questionnaire -

Name and position.

Organisation & department.

Research Topic.

Service received.

Paying or part of Oxford Experimental Information Unit project.

Length of time service has been received.

Profile form.

Profile Compilation Method.

Current awareness habits previous to receiving service.

Changes brought about by service.

Importance of service to user.

Relevance and Recall ratings or estimates.

Output form and comments on it.

Output details comments, (preference for abstracts, keywords etc.)

Up to dateness of service (estimates).

How valuable service is. If it were not free would user want to pay for it and would this be possible and how much the service is worth and what price could be afforded.

Many of comments made by users have been used as comments on the individual systems in the survey of Chapter 2 and in the analysis of Chapter 3 and their attitudes form a basis for the discussion in Chapter 4. A summary of users comments follows.

<u>ASCA. SDC. & CT.</u>

One academic research chemist who was part of the LSE. Evaluation experiment was interviewed. He had a poor opinion of SDC., as his notifications were of little relevance to his research interests though the odd unknown paper had been picked out by this but missed by the others. He felt that CT. and ASCA. were almost equally good but that ASCA.'s citation facility made it especially attractive. He pointed out that SCI. is very out of date when it arrives thus making ASCA. even more valuable. SDI. had cut out his scanning of CT. and reduced his primary journal scanning by half to 20 core journals. Here was a user who was totally convinced that S.D.I. was necessary to him.

<u>INSPEC S.D.I. Investigation.</u>

Two academic research workers in solid state physics were interviewed. A group profile was being run covering themselves and their research students. However, both were lecturers and had very diverse personal interests, thus the profile covered a few topics which were common to their separate projects. These were mainly concerned with the thermal properties of various named semi-conductors. The profile did not fit interests sufficiently to provide a vital service. The service was rated useful. Current Papers in Physics was felt to serve them better. This was up to date enough for document supply available to them. The S.D.I. service was likely to be useful when the user was too busy to do his regular scanning.

Four research and development workers in the electronics side of telecommunications displayed a variety of opinions on the service they were receiving. Again the poor methods of tailoring the profile to the users needs were apparent. In several cases the users were more on the development side and information on pure research in their field was of little interest as development had progressed too far for knowledge of other research to alter the course of action. One of the workers used none of the traditional current awareness type publications and read few journals and his profile was obviously not 'on the ball' probably because the thesaurus could not describe his applied interests specifically enough. Another worker who kept a personal card file had received very few references worth placing in it. His interests were not really covered by the data base, his scanning of journals and current papers publications seemed to him to be more efficient. Another worker was very pleased with his notifications he was getting about 45% which were of some interest. He was interested in new information on integrated circuits and general trends in telephone systems and obviously his interests had been well expressed in his profile. He was pleased with the service as it saved him time. He did not think that it was a vital service as he was not terribly rushed for time and at slack times managed to scan trade publications, he received, fairly adequately. He felt that it would be worth his group receiving such a service as several workers would be well covered by a single profile. He was not at present sharing his S.D.I. output with his co-workers. The fourth worker felt his library should be spoon-feeding him. He was again on the development side and did not feel the need for current awareness: he thought pre-project retrospective searches were sufficient. These research workers received a regular current awareness bulletin from their 'on site' library containing abstracts and most of them felt this service to be adequate.

The Inspec service did not seem really suitable for these research workers; this was mainly because the data base was not suitable for them. Their interests were either not describable by a simple profile of the Inspec form or they did not require a current awareness service anyway. Their profiles did not appear to be well constructed in the first place, users did not appear to understand the system, and the profile modification method obviously did not work very well.

Plasma Physics Current Awareness service.

An academic research worker in Plasma Physics in a small Physics department was interviewed. He had previously worked at the Culham laboratory and had contacts there. His research was more of a spare time occupation in term time as so much time was spent in lecturing; hence his work was not very much affected by other research elsewhere. He got a high relevance in his S.D.I. notifications, perhaps as high as 70%. However, he found difficulty obtaining the documents as the library stocked little in his field, being such a small group interested in that field. Also, when the University library received publications from abroad, these were of the order of months after his notifications and as the output was computer print out and he kept no personal card file, much of his notifications went to waste. He also received the Plasma Physics Current Awareness bulletin free and felt this would serve him as well as S.D.I. The Plasma Physics service was stated to be better than the appropriate section of current papers in physics as it included reports.

MEDLARS

An academic pharmacology research worker who was receiving this service as part of L.S.E. was interviewed. He had been part of the S.C.I.P. and received output from CBAC and his perfected profile was simply converted to the Medlars system convention. The CBAC data base was not really suitable for his interests and Medlars suited better; however the service had been rather unreliable and was so much out of date that he found it useful only as a back up service to his own current awareness system of scanning primary journals and Current Contents Life Sciences. About one third of his notifications were relevant. His interests had fitted in with the thesaurus reasonably well.

The information section of a drug firm dealing with the pharmacology of their products were using Medlars to provide information for their drug files. Medlars was being used as a back up service to their own journal scanning. Ideally the user here would have liked to use drug names only, for retrieval, but many of course were not allowed terms and 'anti-flammatory agents' of parts of the body had to be used instead and inevitably produced a lot of 'Trash'. Relevancy of about one third was achieved. CBAC was also used for this purpose. This provided a lower output but higher relevance.

A group of academic ergonomics workers were interviewed. They were in the process of compiling a profile to cover their entire interests and were finding this very difficult both because these were fairly diverse and because the thesaurus terms could not adequately describe their interests.

UKCIS

Three final year postgraduate students were interviewed who were running profiles on C.T. as part of the L.S.E.

One was a physical chemist interested in flash photolysis. He received a very large output usually more than one hundred items per notification. This was felt to be satisfactory as the print out took little time to scan and was quicker than scanning C.C.P. and primary journals. The service covered more journals than the library took and helped with foreign publications. The service was felt to be useful as it left more time for reading the actual articles.

An inorganic chemist studying the chemistry of specific compounds had an average output of around thirty items of which two or three were important; a fair amount of trash had appeared from the use of the named compounds in the medical and agricultural fields. This student relied on the service entirely after initially scanning all the major primary journals as well and finding nothing in addition. This user felt his topic straightforward enough to be covered by just scanning journals.

The third profile was run for a group of electrochemistry workers who had been running it previously as part of SCIP. They found it useful and took it for granted though still scanned major primary journals. They intended to continue receiving it and were confident they could find financiers from industry.

An industrial organisation interested in coal by-products were subscribers to C.T. and also were part of the C.A.C. experiment. A broad profile was being run by the information department who disseminated individual items to likely staff and used the output as an information source for their current awareness bulletin. The C.T. service had been run since the early days of C.S.R.U. and was running steadily now. The C.A.C. service was producing large outputs and was rather unreliable.

The chemical research workers in a drug firm were running four profiles on CBAC for different group interests. They had previously used C.T. but in their field titles were not much use. Pure research forms only a small part of their interests and patents are their prime interest, as far as current awareness is concerned and for this Derwent publications are used. S.D.I. took third place in their information system with an 'on site' library produc~~ing~~ed current contents bulletin taking second place to patents information. The small coverage of CBAC means that C.A. still has to be used. Perhaps C.A.C. will provide the answer to the user.

Users were not found to be particularly impressed by the S.D.I. services they received, though many found them useful to a greater or lesser extent. Improvements in some services, better user education and more care in profile construction and modification would probably make them more attractive to the user. Users who understood the services they were receiving got more from them.

APPENDIX III

Machine Readable Data Bases. (29)

	Cost.	Tape Users.
Chemical Abstracts Service		
C.T.	£717 (excluding cost of actual tape)	Aldermaston, Shell U.K.C.I.S., Denmark.
CB/C	£738 (" " " " ")	U.K.C.I.S.
POST - J	£708 (" " " " ")	U.K.C.I.S.
POST - P	£695 (" " " " ")	U.K.C.I.S.
C.A.C.	£1,685 (" " " " ")	U.K.C.I.S.

(Minimum software is supplied with trial files of these.)

	Cost.	Tape Users.
Biological Abstracts		
Biological Abstracts Previews		U.K.C.I.S.
Engineering Index		
Electrical/ Electronics	$1,600-2,000	European tests.
Plastics	$1,400-1,800	Eastman/Kodak U.S.A.
INSPEC		
S.D.I. Experiment tapes		Aldermaston.
Full Inspec tapes (Jan.1970 onwards)	Whole of Science Abstracts £2,500	
	Sci.Abs. without abstracts £1,750	(provisional estimates)

	Cost.	Tape Users.
<u>INSPEC</u>(cont.)		
	Sections of Sci.Abs. with abs. £1,750	
<u>I.S.I.</u>		
Source Tapes	$6,000-8,000	Unilevor.
Source Tapes + +Citations+patents	c.$26,000	
Index Chemicus	$7,000	L.S.E. with U.K.C.I.S.
<u>MEDLARS</u>		N.L.L.
<u>Nuclear Science Abstracts</u>		
N.S.A.	(Not generally available. Harwell and Aldermaston have them on an exchange basis)	
<u>N.A.S.A.</u>		
S.T.A.R. I.A.A.	(Not generally available, on exchange agreement with ESRO/ELDO)	
<u>PANDEX</u>	$6,500	U.S.A. custom service.

APPENDIX IV

S.D.I. Software packages (29) (77)

a) Commercially available packages:

Chemical Abstracts Service made software packages available with their data tapes. Originally these were for the IBM 1401 series and were used by Eli Lilly and Olin Mathieson Companies in the early days of C.T. Now a minimum software package is provided with the data base for new subscribers for trial runs: this is for the IBM 360 series. Aldermaston and Shell used the C.A.S. packages to start with but had problems.

Ames Laboratory devised their own software system which is very sophisticated and includes automatic profile updating. This is flexible and will take I.S.I., C.T. and N.S.A. input formats. This is a well tried system but probably more complex than required by many users. It is written in COBOL and requires a core store of 200K bytes with four tape files, one data disc file and four 1,500 track discs for sorting. This is suitable for the IBM 360 series.

National Research Council of Canada programs were written for IBM 360 machines and require less than 256K bytes core store. The programs are fairly simple and were originally written for C.T. tapes. Boolean logic, term truncation and weighting is used. Unilever and Aldermaston are using these programs. It is apparently fairly easy to use a conversion program with other tape data bases to make them into the acceptable C.T. format.

<u>I.B.M. ITIRC/TEXPAK</u> is the system used by IBM for their 'in house' S.D.I. system and involves a complicated input form not immediately suitable for machine readable data base tapes but more for internally keypunched input. This system is designed for the IBM 360 series.

<u>I.B.M. Document Processing System</u>

This is part of IBM's G.I.S. (General Information System) designed for normal text retrospective searching. This is very flexible as far as input format is concerned and incorporates good search logic. This is not designed for S.D.I. but could probably be modified for such use.

<u>I.C.S. Ltd. Kidsgrove</u> (6) has a package designed for KDF9. This is designed for a thesaurus based system and allows some flexibility in input format. This package is somewhat similar to that used in the INSPEC S.D.I. investigation.

b) <u>'In house' system software</u>

<u>U.K.C.I.S. Nottingham.</u> Based on C.A.S. programs and experienced C.S.R.U. wrote their own programs for the KDF9 at Nottingham University which has peripherals not in the same series. These programs accept the different formats of the data bases used.

MEDLARS N.L.L. search programs were specially written for the Newcastle University KDF9 computer. The University of Colorado runs its Medlars searches on a Honeywell computer. The University of California uses an IBM computer.

Culham wrote their own program in virtually non-revisable user code for their KDF9 computer. New programs have been written for the E.E. 470. Culham uses internally keypunched input.

GKN wrote their own programs for IBM 360/30 in PLI and Assembler languages for a thesaurus controlled internally keypunched input.

Harwell wrote their own programs for IBM 360/65 in FORTRAN and Assembler languages for N.S.A. tape input.

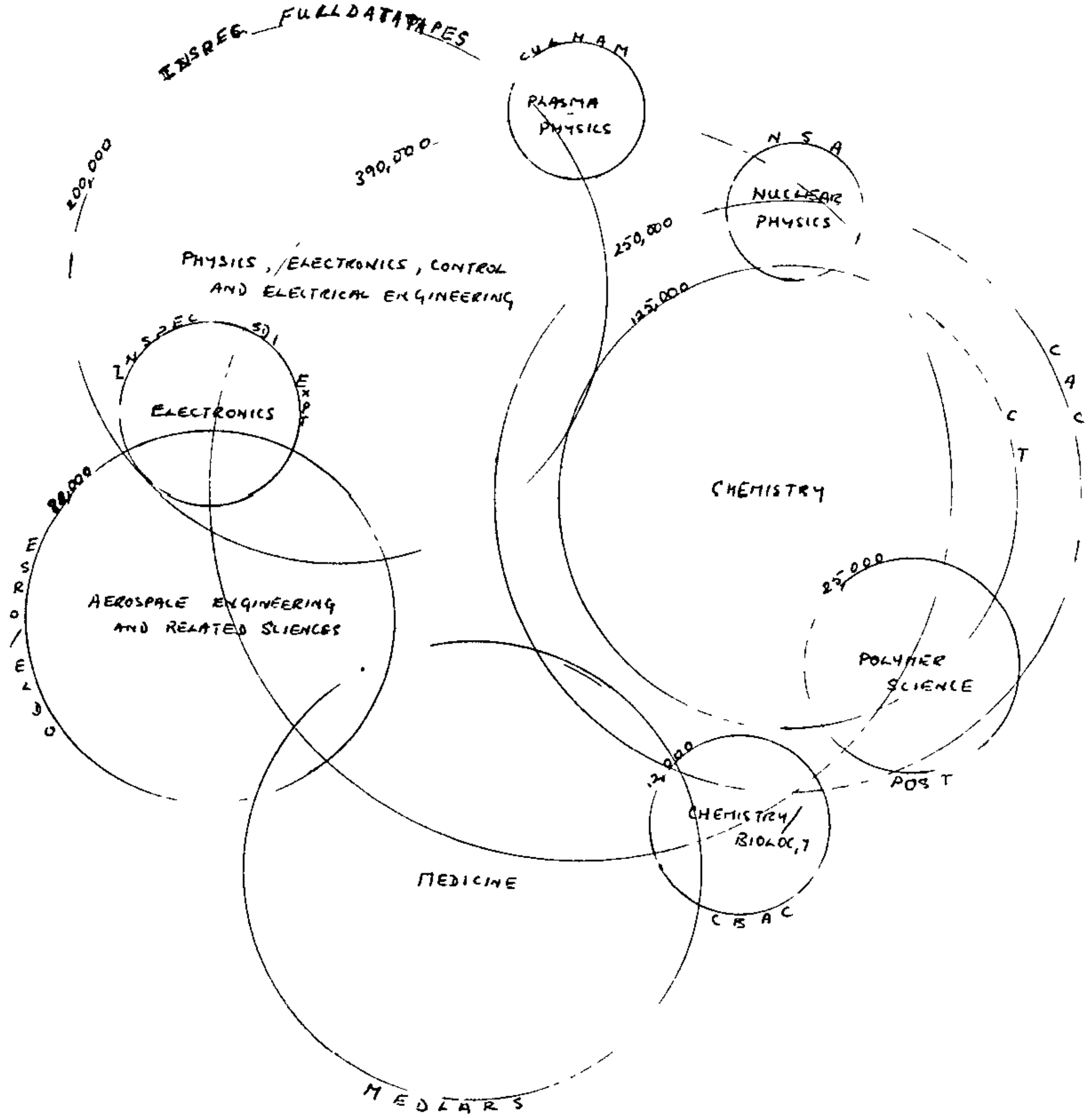

Fig. 1 Comparative input volumes and subject coverage overlap of S. D. I. services

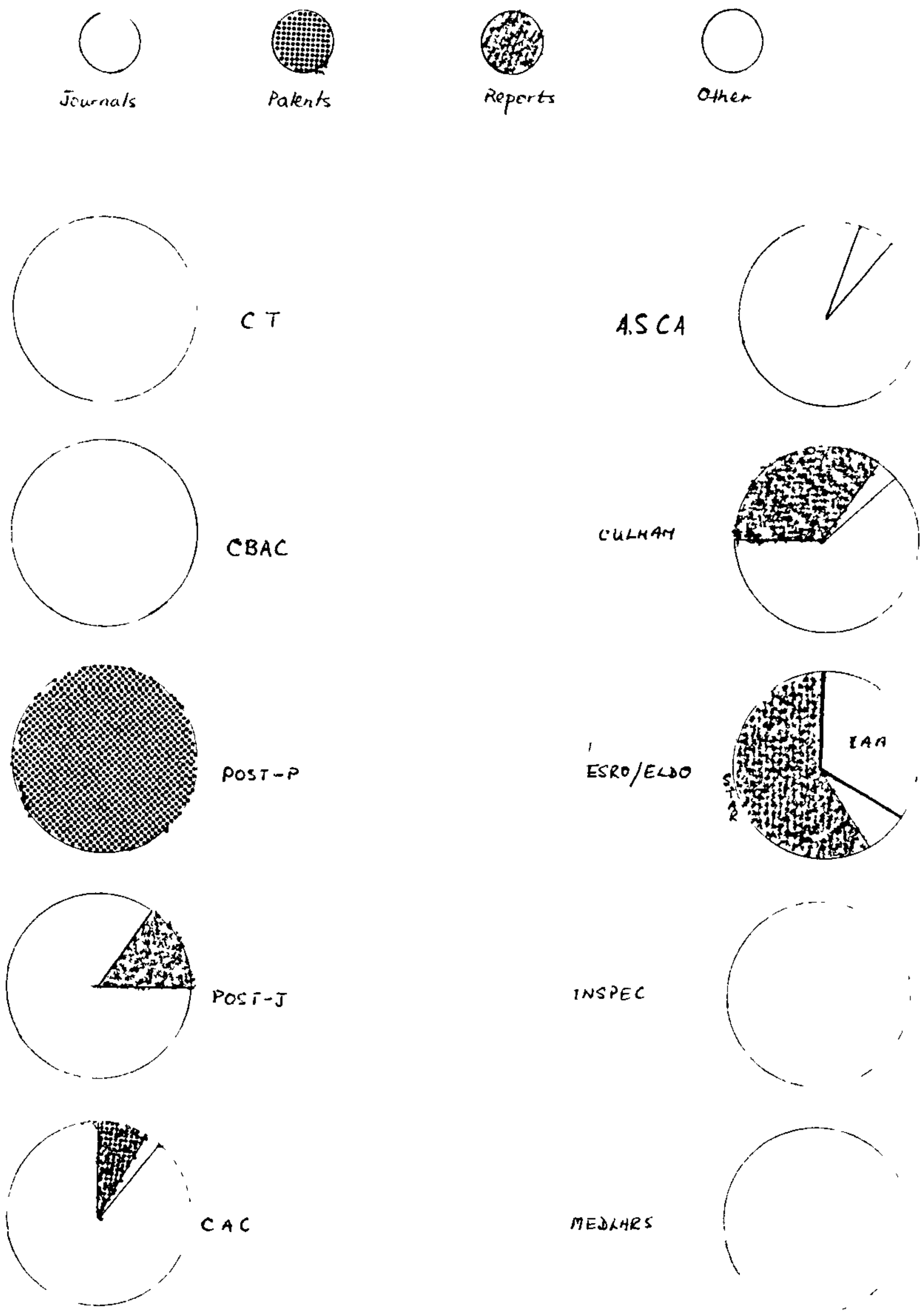

Fig. 2 Form coverage of custom S.D.I. services.

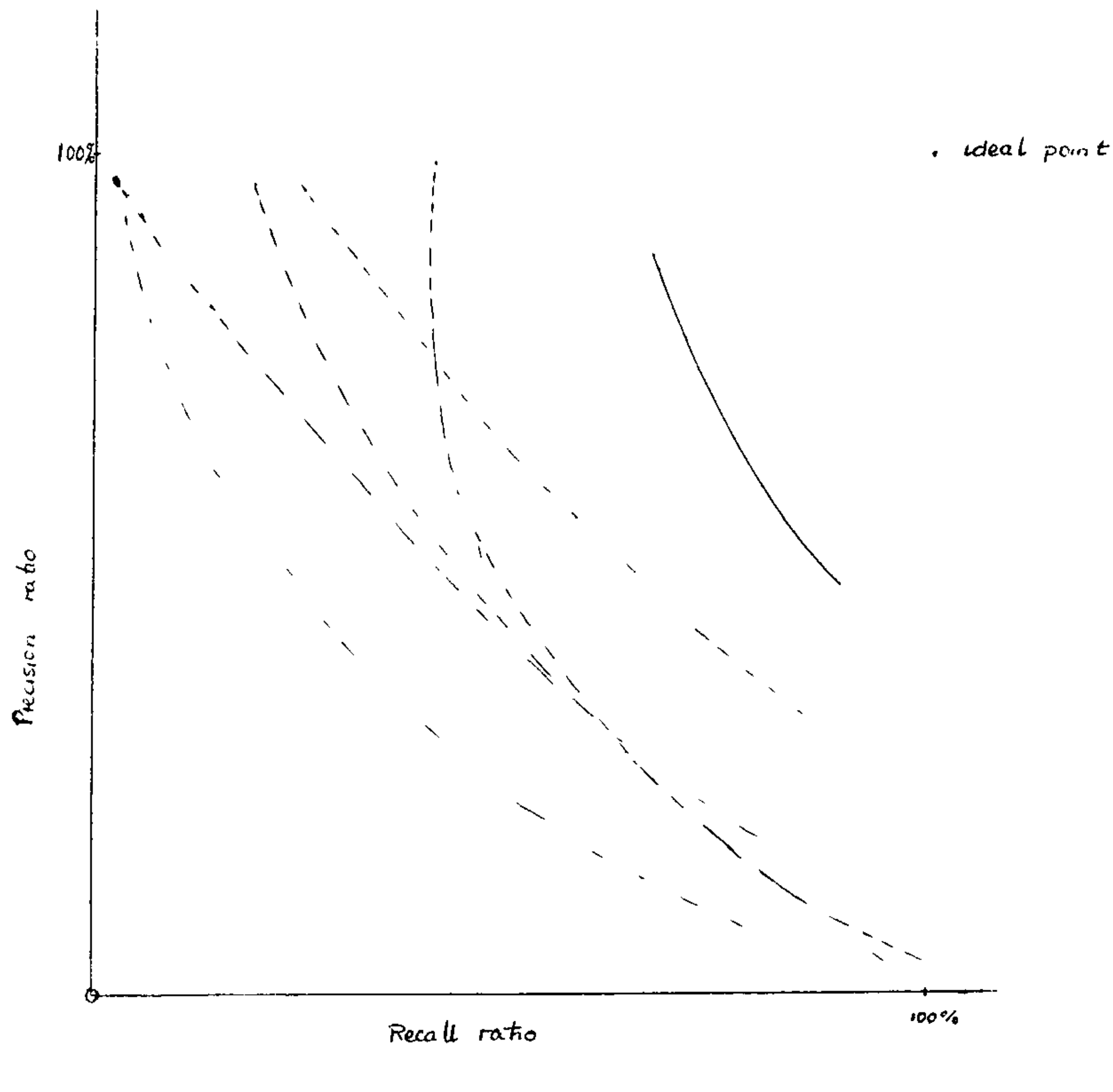

Fig. 3 Operational curves for hypothetical S. D. I. systems.

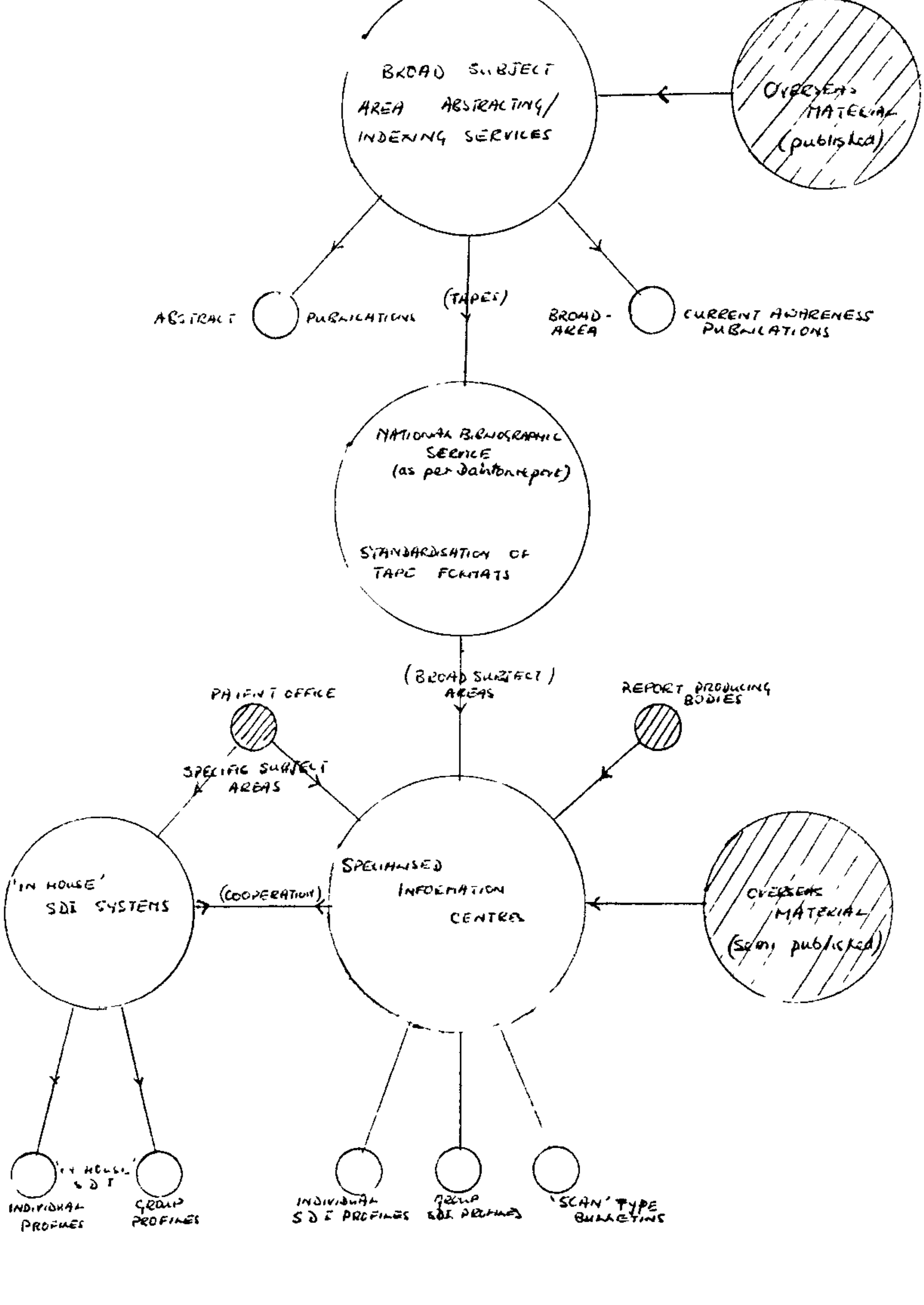

Fig. 4 National Information Network of the future

Current Awareness practises in Industry

Please answer all the main questions by marking the appropriate box and the question parts where the answer to the main question is 'yes'. Please add extra details where applicable particularly for question 5.

Name of organisation ....

How are your research and development scientists and technologists kept abreast of current developments in their subject fields? -

1. Are they expected to scan primary journals in the library? YES NO
2. Are current journals in their fields circulated to them? YES NO
3. Are they expected to scan abstract/current awareness journals in the library? YES NO
4. Is a current awareness bulletin issued by the library? YES NO

   (a) Who issues this bulletin? the central library / the on site library

   (B) Is it subject divided? YES NO

   (c) What does it include? individual journal article refs / just library accessions / internal reports

5. Is some form of selective dissemination of information service* provided? YES NO

   (a) Who to? Just group leaders / All qualified research staff

   (b) How many users? ....

   (c) Is it a manual system obtained by library staff scanning incoming documents? YES NO

   (d) Is it a mechanised system? YES NO

   (i) Are profiles run on an external SDI service (egA.S.C.A.) YES / NO

   How many?..............Which services?..............

   (ii) Is the SDI service provided by the library/information unit? YES / NO

   What form of input is used?

   Key punching of document details by library/information staff

   Commercial magnetic tapes(eg C.T.,CBAC,ASCA)

* Selective dissemination of information is taken here to be a personal listing of document references likely to be of interest to a user, obtained by matching user interest profiles with incoming documents.

THANK YOU VERY MUCH FOR YOUR CO-OPERATION