EUR 5169 e

COMMISSION OF THE EUROPEAN COMMUNITIES

USER REACTIONS TO "CAIN"

(Cataloguing and Indexing Data Base of the "Bibliography of Agriculture")

by

L.H. CAMPEY

1974

Prepared for consideration by the
Working Group for Agricultural Documentation and Information
of the European Communities

LEGAL NOTICE

This report is on sale at the addresses listed on cover page 4

at the price of B.Fr. 70.—

Commission of the
European Communities
D.G. XIII - C.I.D.
29, rue Aldringen
L u x e m b o u r g

July 1974

EUR 5169 e

USER REACTIONS TO ,,CAIN'' (Cataloguing and Indexing Data Base of the ''Bibliography of Agriculture'') by L.H. CAMPEY

Commission of the European Communities
Prepared for consideration by the
Working Group for Agricultural Documentation and Information
of the European Communities
Luxembourg, July 1974 - 56 Pages - B.Fr. 70.—

The results of discussions and correspondence with ten out of twelve known subscribers to CAIN, including the producers of the Bibliography of Agriculture are described. They cover both the performance of the data base and user reactions to it and they are discussed under the following headings- coverage, accessibility, suitability for selective dissemination of information and suitability for retrospective searches. The report concludes that CAIN provides good coverage of the agricultural literature, including a substantial amount of useful fringe material, but that it must be supplemented by alternative services for complete coverage. Items in CAIN are frequently

—

EUR 5169 e

USER REACTIONS TO ,,CAIN'' (Cataloguing and Indexing Data Base of the ''Bibliography of Agriculture'') by L.H. CAMPEY

Commission of the European Communities
Prepared for consideration by the
Working Group for Agricultural Documentation and Information
of the European Communities
Luxembourg, July 1974 - 56 Pages - B.Fr. 70.—

The results of discussions and correspondence with ten out of twelve known subscribers to CAIN, including the producers of the Bibliography of Agriculture are described. They cover both the performance of the data base and user reactions to it and they are discussed under the following headings- coverage, accessibility, suitability for selective dissemination of information and suitability for retrospective searches. The report concludes that CAIN provides good coverage of the agricultural literature, including a substantial amount of useful fringe material, but that it must be supplemented by alternative services for complete coverage. Items in CAIN are frequently

—

EUR 5169 e

USER REACTIONS TO ,,CAIN'' (Cataloguing and Indexing Data Base of the ''Bibliography of Agriculture'') by L.H. CAMPEY

Commission of the European Communities
Prepared for consideration by the
Working Group for Agricultural Documentation and Information
of the European Communities
Luxembourg, July 1974 - 56 Pages - B.Fr. 70.—

The results of discussions and correspondence with ten out of twelve known subscribers to CAIN, including the producers of the Bibliography of Agriculture are described. They cover both the performance of the data base and user reactions to it and they are discussed under the following headings- coverage, accessibility, suitability for selective dissemination of information and suitability for retrospective searches. The report concludes that CAIN provides good coverage of the agricultural literature, including a substantial amount of useful fringe material, but that it must be supplemented by alternative services for complete coverage. Items in CAIN are frequently

difficult to access, so that the formulation of search profiles normally requires considerable skill and effort. With these reservations CAIN can give acceptable results on SDI and retrospective searches when professional assistance in profile construction is available. The report also includes general observations on the implications of these results for AGRIS.

difficult to access, so that the formulation of search profiles normally requires considerable skill and effort. With these reservations CAIN can give acceptable results on SDI and retrospective searches when professional assistance in profile construction is available. The report also includes general observations on the implications of these results for AGRIS.

difficult to access, so that the formulation of search profiles normally requires considerable skill and effort. With these reservations CAIN can give acceptable results on SDI and retrospective searches when professional assistance in profile construction is available. The report also includes general observations on the implications of these results for AGRIS.

EUR 5169 e

COMMISSION OF THE EUROPEAN COMMUNITIES

USER REACTIONS TO "CAIN"

(Cataloguing and Indexing Data Base of the "Bibliography of Agriculture")

by

L.H. CAMPEY

1974

Prepared for consideration by the
Working Group for Agricultural Documentation and Information
of the European Communities

ABSTRACT

The results of discussions and correspondence with ten out of twelve known subscribers to CAIN, including the producers of the Bibliography of Agriculture are described. They cover both the performance of the data base and user reactions to it and they are discussed under the following headings- coverage, accessibility, suitability for selective dissemination of information and suitability for retrospective searches. The report concludes that CAIN provides good coverage of the agricultural literature, including a substantial amount of useful fringe material, but that it must be supplemented by alternative services for complete coverage. Items in CAIN are frequently difficult to access, so that the formulation of search profiles normally requires considerable skill and effort. With these reservations CAIN can give acceptable results on SDI and retrospective searches when professional assistance in profile construction is available. The report also includes general observations on the implications of these results for AGRIS.

TABLE OF CONTENTS

****

PREFACE

This report has been prepared to help those interested in the European use of the FAO AGRIS I information system to judge the possible value of a bibliographic service based on its magnetic tapes. AGRIS I will resemble the existing CAIN system, developed by the US National Agricultural Library, in that it will be designed for search by "natural-language" or "free-text" methods and not by recourse to a thesaurus of indexing and retrieval terms. Experience with the CAIN system is therefore a useful indication of what the AGRIS I system can achieve in practice.

A dozen organisations in different countries have been experimenting with CAIN tapes and, although the projects have been designed independently with no common purpose or methodology, they have had sufficient points of similarity to justify an analysis of the collective experience. After discussion of this possibility in the Working Group for Agricultural Documentation and Information of the European Communities, two national organisations, the Institut für Dokumentationswesen and the Office for Scientific and Technical Information (now part of the Research and Development Department of the new British Library) offered to finance independently a systematic study. Mrs. L.H. CAMPEY was engaged to conduct the study. She worked entirely on her own, receiving various reports on the projects and visiting most of the responsible organisations. Her methodology is described in the introduction to her report.

In the view of the Working Group the results of this study have fully justified the effort; and since they are of wide interest to agricultural and information specialists, the sponsoring bodies have gladly accepted an offer by the Commission of the European Communities to publish the report.

Section 1. INTRODUCTION

This report has been prepared at the request of the UK Office for Scientific and Technical Information (OSTI). A previous study ("A comparative assessment of user reactions to CAIN", May 1973) drew conclusions about user search experience with the CAIN tapes on the basis of an evaluation of published reports and private correspondence. Subsequent to this, OSTI commissioned a further study with the following terms of reference:

"To carry further the study on recent investigations of the use of the CAIN tapes with the aim of obtaining by discussions with users of CAIN tapes as much additional information as possible on the performance of the system and user reaction to it. These discussions should be held with the AGRIS system in mind so that the results of the study can be presented in terms of the implications for AGRIS if required."

In the preparation of this report, a number of CAIN subscribers in both the USA and Europe were visited. Discussions were held with the person responsible for providing an information service, either based exclusively on CAIN or which included CAIN as well as other tape services. In addition, further evidence of search experience with CAIN has been obtained from reports and letters and these findings are also presented in this report.

Data has been collected for 11 out of a total of twelve known subscribers. It must be remembered that one of these subscribers - MacMillan Publishing Co., uses CAIN for producing the Bibliography of Agriculture (B of A) and information relating to the B of A tape service is discussed in a separate section of the report (Appendix A). Data obtained for the remaining ten subscribers forms the basis of this report, together with the findings provided by a single "one-time only" user of the CAIN tapes. In all, the report presents and summarises the information collected from visits to six subscribers and private correspondence and reports received from four subscribers and one additional CAIN user.

The remainder of this report is divided into three major sections. The first of these (Section 2) contains a brief description of the CAIN tape service while in Section 3, the search experience and reactions to CAIN reported for eleven users are summarised. This summary is supported by a more detailed account of user search experience with CAIN and this appears as a separate section (Appendix B). Based on the main findings of the study, a number of conclusions have been drawn relating to CAIN's search performance and user reactions to it, and these appear in Section 4, the final section in the main body of the report. Some general observations based on the conclusions arrived at from this study, which may have implications for AGRIS, are also presented in Section 4.

Section 2. DESCRIPTION OF CAIN (CAtaloguing and INdexing)

2.1 SUPPLIER: National Agricultural Library, U.S. Department of Agriculture (USDA), Beltsville, Maryland 20705.

2.2 SUBJECT COVERAGE: CAIN is a monthly tape service containing bibliographic data on documents acquired by the National Agricultural Library (NAL) in the broad field of agriculture, including: agricultural and rural sociology, agricultural products, animal husbandry, engineering, entomology, food and human nutrition, forestry, pesticides, plant science, soils and fertilizers, and other related subject fields. About 90% of the references contained in the data base are journal articles and other input sources include: monographs, series publications, government reports including USDA literature, experimental station reports and extension service publications.

In addition to the monthly service initiated in 1970, CAIN also incorporates a separate pesticides file containing 42,667 items collected in the period from 1967 to 1969. Including this pesticide material, the total cumulative CAIN file is reported to contain 479,354 records as of August, 1973, according to recent figures supplied by the NAL. Statistics giving the number of items included in each yearly volume have been provided by the University of Georgia and these are reported in Section B.4.2 of this report.

A comparative study of the subject coverage provided by CAIN relative to other agricultural and agriculturally-related data bases is currently being carried out by the CAIN/CAB Project Group at Nottingham. Progress made thus far in this investigation is outlined briefly in Section B.10.

2.3 NATURE OF THE SERVICE: NAL prepare the CAIN tapes and use them for various internal applications such as producing the NAL card and book form catalogues and printing numerous reports and bibliographies. The CAIN tape service was first made available in January 1970 to outside users and no associated search software was supplied with the tapes. It was the first major agricultural machine-readable data base to be released and there are currently twelve subscribers. One subscriber, Macmillan Information, uses the CAIN tapes to produce printed and magnetic tape versions of the Bibliography of Agriculture (B of A). A description of the B of A tape service together with the names of four subscribers appear in Appendix A. A brief account of search experience reported by two of these B of A users is also included.

The CAIN tapes are supplied by NAL on a monthly basis at a cost of $45.00 per reel.

2.4 CONTENTS OF TAPE RECORDS: Only some portions of the CAIN tape records are applicable to the search requirements of subscribers. The remaining data is intended for different internal NAL functions and will not be considered in this report. The important searchable data elements in each record are:

a. Subject elements

i Title - always with English translation if required and about 10 to 15% are enriched,

ii Subject category codes - a maximum of two per item selected from a total of 70 subject classifications (about 10% of items are assigned more than one category code),

iii Subject terms - assigned only to monographs at present and an average of 1.5 uncontrolled subject terms are presently being assigned by indexers.

b. Author and citation elements such as:

i personal and corporate authors,

ii journal title,

iii document type and language and availability of a translation.

Analyses of the distribution of CAIN items under the different subject category codes have been carried out in Sweden and Norway. Their findings are reported in Appendix C.

2.5 CAIN ON-LINE SERVICE: The award of a contract to Lockheed Missiles and Space Company, Inc. to provide an on-line interactive bibliographic search and retrieval service for CAIN was announced in June of this year. This system is intended to serve primarily the bibliographic information needs of the U.S. Department of Agriculture through the Library. The service will also be available to anyone in the agricultural community at a modest cost.

The bibliographic data for the system will be supplied by NAL while Lockheed are to take charge of operating the on-line service, which will provide retrospective searches of a cumulative CAIN file that includes all records from January 1972. The CAIN on-line service will be available to users at a cost of $25.00 per hour, exclusive of communication line charges.

No details of any search experience with this new on-line system have been obtained.

Section 3. SUMMARY OF THE MAIN FINDINGS

Detailed information of user search experience with the CAIN tapes obtained as a result of discussions with users and from a review of published reports and private correspondence is contained in Appendix B. An attempt has been made to extract the most significant findings and these will now be presented.

In the first part of the summary, information collected from visits to six CAIN subscribers is reported under the following headings:

(i) CAIN search experience - details of SDI and retrospective searches run against CAIN.

(ii) Quantitative studies of CAIN - results of any formal monitoring of CAIN's search performance or user reactions to it.

(iii) User assessments of CAIN's strengths and faults - user views which in some cases are based on subjective assessments rather than on quantitative evidence of CAIN's benefits or failings.

(iv) Overall user reaction - statement of user satisfaction or dissatisfaction and principal reasons for this reaction.

The second and remaining part of the summary records the search experience of five additional CAIN users for whom data has been obtained mainly by letter and from published reports.

(a) Data collected from six user visits

3.1 USER - Lantbrukshogskolan, Ultunabiblioteket, Uppsala, Sweden

3.1.1 CAIN search experience - The Library of the Agricultural College in Uppsala have been operating an SDI service based exclusively on CAIN since January 1971 for scientists at the College and at other organisations throughout Sweden. The CAIN SDI service was offered free of charge to users throughout the first 24 months of operation, but since January 1973 it has been running on a cost recovery basis. Prior to the introduction of charges, there were 119 CAIN profiles, and immediately after the charging system was adopted the numbers dropped by 40%. As of August, 1973 a total of 86 CAIN profiles, representing 70 users, were being processed by the Library.

All user profiles are constructed at the Library by an information specialist who throughout the entire period of operation has formally monitored search performance as expressed by the relevance of the search results. CAIN searches are run at the FOA Index (a section of the Research Institute of Swedish National Defence) who provide the necessary software and machine facilities for execution of the various search runs. The search system used by FOA Index is a general purpose IR package which provides as one of its functions, the searching of commercial tapes such as CAIN.

3.1.2 Quantitative studies of CAIN - The following studies have been carried out by the Swedish Agricultural College:

(i) 1971-1972 relevance data - An analysis of user relevance assessments for this two year period has been completed. The results show that around 60% of the hits obtained were of prime importance and 10% were of secondary interest. Including items classified as being of secondary interest, an average relevance of 70% has been demonstrated for the search results. One other noteworthy point to emerge from this study is that only 4% of the total number of titles printed as hits gave inadequate information for the purpose of assessing relevance.

(ii) 1973 relevance data - detailed studies of the percentage relevance of search results obtained for individual profiles are currently being carried out. The preliminary findings indicate that about half of the total number of profiles studied thus far are giving search results, in which 60% or more of the hits obtained are relevant to the question posed.

3.1.3 User assessments of CAIN's strengths and faults - CAIN's chief strengths are considered to be its good broad coverage of the agricultural literature and its relatively inexpensive subscription charges. It is thought to be particularly strong in its coverage of East European literature and the more obscure publications that are not normally seen by users. However, the Library feel that CAIN's fairly extensive coverage of non-scientific advisory journals can become a source of irritation to research scientists when these items appear frequently in their search results.

Like most users visited, Lantbrukshogskolan consider CAIN's principal weakness to be its inadequate representation of the subject content of items covered by the service. In their view, the most effective method of improving CAIN in this area would be to assign more subject category codes to documents without necessarily extending the total number beyond 70. As a secondary improvement, they consider that the addition of more detailed subject terms to enhance weak titles would be beneficial in enabling the retrieval of relevant documents which might otherwise have failed to appear as hits.

3.1.4 Overall user reactions - Users are generally very satisfied with CAIN and rely on it to a large extent for their current awareness needs, even though they must pay for the service. However, users consider that CAIN needs to be supplemented by other sources if they are to achieve complete coverage of their subject fields.

3.2 USER - University of Florida, Institute of Food and Agricultural Sciences (IFAS), Gainesville, Florida

3.2.1 CAIN search experience - Hume Library currently run an SDI service based only on CAIN for scientists employed by IFAS. The service is in its fourth year of operation and 220 profiles are now being searched at no charge to the users. The number of profiles searched has more or less reached a constant level by now after rising sharply during the initial period of operation, but the trend has always been toward increased usage of the service. The Library have not formally monitored the relevance of the search results but are of the general impression that about 70% of the total number of hits obtained are relevant to the questions posed.

The University of Florida have not had any experience in running retrospective searches of CAIN, but they are planning to use the Lockheed on-line service in the near future (see Section 2.5).

All user profiles are prepared by the Library staff who regularly monitor overall user reactions to the search output received. The CAIN search runs are carried out by the University Computer Department using software specifically designed for this SDI application.

3.2.2 Quantitative studies of CAIN - One study has been completed recently by the University of Florida:

(i) Cost/benefit study - In 1973, the Library undertook a questionnaire survey of user reactions to CAIN in order to obtain some quantitative evidence of benefits that were being derived from the service. The results indicate that the majority of users (80%) find the service to be of material value, particularly in terms of hours of research time saved. It has also been demonstrated that the cost savings that result from a reduction of research time spent in current awareness activities are far greater than the total human and machine costs of operating the CAIN SDI service.

3.2.3 User assessment of CAIN's strengths and faults - Although they are generally very satisfied with CAIN, the University of Florida did comment on two areas in which they consider CAIN to be weak. Firstly, they consider that more enrichment terms should be added to titles to give a better representation of the important subjects contained in a document, and secondly, they would like to see an improvement in the coverage provided in the food science and agricultural economics subject fields. CAIN's chief strengths are considered to be its good broad coverage of the agricultural literature and its low acquisition costs.

3.2.4 Overall user reactions - On the whole, IFAS users are very satisfied with their CAINsearch results. Users with access to library facilities normally check other sources to supplement these results, while scientists at remote experimental stations throughout Florida rely solely on CAIN for their current awareness needs.

3.3 USER - Centre for Agricultural Publishing and Documentation (PUDOC), Wageningen, the Netherlands.

3.3.1 CAIN search experience - PUDOC provide as one of their functions a national information service for agricultural literature. They are currently evaluating the usefulness of a computer-based information service in satisfying the current awareness needs of Dutch agricultural scientists. Seven commercial tape services including CAIN and the Bibliography of Agriculture (B of A) are being tested on an experimental basis, with users reporting back on the relevance of hits received to allow an assessment of the performance of the various tapes searched.

PUDOC began using CAIN in September 1972, and are currently running 54 SDI profiles against CAIN and different combinations of other tape services at no charge to users. Each profile is searched against CAIN and B of A and some are also searched against the other available tapes, depending on the subject content of the profile.

All SDI profiles are constructed by an information specialist who provides the interface between the user and the different tape services. PUDOC run CAIN, B of A and Food Science and Technology Abstracts as in-house services in co-operation with the Institute for Mathematics, Information Processing and Statistics, the Hague, who supply the necessary software and machine facilities for execution of the search runs. PUDOC have also made arrangements with two other organisations to gain access to four additional tape services on a service bureau basis.

3.3.2 Quantitative studies of CAIN - The following three studies have been carried out by PUDOC:

(i) Relevance studies - The relevance data obtained for nine SDI profiles tested over a period of nearly one year indicated that CAIN and B of A performed slightly better than the other tape services when viewed in the context of the total number of profiles analysed. However, in cases where a profile fell into the scope of "non-agricultural" data bases such as MEDLARS and Chemical Abstracts Condensates, these tapes often out-performed CAIN in terms of the percentage of relevant hits provided.

(ii) Relevance/recall studies - The relevance data collected for the 54 profiles that are currently being run, indicate that a relatively high percentage of the hits retrieved from CAIN are relevant to the particular questions posed. Comparisons of the relative recall obtained for the different tape services are being carried out at present. Very little data is available thus far in the study, but preliminary findings point to the need for a multi data base approach if complete coverage of the agricultural literature is to be achieved. The results show that while CAIN misses relevant hits that other tape services provide, it also retrieves relevant hits that are not found by searching the other tapes.

(iii) Comparisons of CAIN vs. B of A - PUDOC have recently completed a study of the main differences between CAIN and B of A and their effect on search performance. Based on the results of this investigation, PUDOC consider CAIN to be the preferred data base, mainly because of its inclusion of monographs and the foreign language version of titles which are not provided in B of A.

3.3.3 User assessments of CAIN's strengths and faults - PUDOC consider that CAIN provides good broad coverage of the agricultural literature in that it contains many articles of major importance as well as useful fringe material. This view has been confirmed by a study of the coverage of Dutch publications which demonstrated that a high proportion of this material was being included in CAIN. The speed of coverage (i.e. currency) achieved by CAIN is thought to be good and principal strengths are considered to be its coverage of East European literature and the more obscure publications not normally seen by users. Although PUDOC users consider that CAIN titles are adequate for assessing the relevance of hits, they would prefer to see the titles supplemented by subject descriptors and further category codes to allow items to be more accessible during searches.

3.3.4 Overall user reactions - PUDOC users are generally satisfied with the CAIN service and consider that it provides good broad coverage of the agricultural literature. However, their studies indicate that for comprehensive coverage of the agricultural literature, CAIN must be supplemented by other tape services which cover agriculture within other scientific disciplines.

3.4 USER - University of Georgia, Office of Computing Activities, Athens, Georgia.

3.4.1 CAIN search experience - The Information Services Division of the Office of Computing Activities have been offering computer-

based information services to scientists within the 30 institutions in the University System of Georgia as well as other academic, government and commercial organisations for a number of years. The Centre is presently searching 20 commercial tape services for a total of 1661 users where users are defined as mailing addresses. Over 9000 SDI profiles are currently being searched and a total of 6624 retrospective searches have been run by the Centre from June, 1972 to June, 1973. The service is run free of charge to scientists within the University System and on a cost recovery basis to outside users.

The Georgia Centre started its use of CAIN in 1970 and by June, 1971 were processing 196 profiles against CAIN. In 1972, the number of profiles increased to 395 and as of August, 1973, the Centre was searching 601 CAIN profiles. As can be seen from these figures, the growth in the usage of CAIN for SDI has been at a rate of roughly 200 profiles per year.

During a one-year period from June 1971 to June 1972, the Centre ran 224 retrospective searches of CAIN, and during the following year (1972-73) the number of searches increased to about 400. In August 1973, the Georgia Centre ran 37 retrospecitive queries against CAIN, which if extended to a yearly rate would imply that the total number of searches in the next year is likely to be at least 500. Retrospective searches of CAIN are run at two-weekly intervals, and the batch size currently averages at about 20 queries.

In most cases, profiles are prepared by an information specialist who is familiar with the indexing and editorial policies followed by the different data base suppliers. The search software has been designed to provide efficient processing of a wide range of large, bibliographic data bases.

3.4.2 Quantitative studies of CAIN - No formal monitoring of the performance of any of the tape services or user reactions to them has been attempted.

3.4.3 User assessments of CAIN's strengths and faults - CAIN's principal strengths are considered to be its good broad coverage of the agricultural literature, its low acquisition costs, and the fact that relatively low machine running costs are incurred for SDI and retrospective searches. Although they feel that CAIN titles are adequate for assessing the relevance of hits, they would like to see these titles supplemented by more subject descriptors for search purposes. In addition, the Centre consider that the subject category code system should be extended and that more codes should be assigned to each item in CAIN. However, it must be remembered that these views do not reflect any quantitative evidence of CAIN's merits or failings in these specific areas, but are merely general impressions based on experience gained in operating the information service.

3.4.4 Overall user reactions - The Georgia Centre was reluctant to commit itself to an opinion as to whether CAIN users were satisfied or not since there is virtually no quantitative evidence to support any firm views on the matter. The general impression is that CAIN does give satisfactory results for agricultural scientists working in the applied fields or in cases where profiles do not require extensive coverage of peripheral subject areas. For many profiles, CAIN is combined with other tapes such as Biological Abstracts, to provide full coverage of the user's subject field but there

are some cases when profiles are searched only against CAIN. It is felt that since the number of new CAIN searches is increasing at a rate which is close to the overall average rate for all data bases, it can be assumed that CAIN users are at least as satisfied as users of many of the other tape services.

3.5 USER - Agricultural Research Service, United States Department of Agriculture (USDA), Beltsville, Maryland.

3.5.1 CAIN search experience - The Agricultural Research Service (ARS) have been operating an SDI service free of charge to USDA scientists since May 1972. The service incorporates CAIN and four other commercial bibliographic data bases, and searches are provided for a user community of just over 1,000, for whom around 6,000 SDI profiles are currently being run. CAIN is not the dominant data base as far as usage is concerned, with there being only 600 CAIN users compared with about 1,000 users of Chemical Abstracts Condensates (CAC) and 800 users of Biological Abstracts.

A total of just over 1,000 SDI profiles are presently being searched against CAIN on behalf of these 600 users. As a general policy, CAIN is always searched in combination with other tape services such as CAC and Biological Abstracts, since it is felt that CAIN on its own does not give comprehensive coverage of the user's field of interest. However, it is also generally recognised that these other tape services are likely to be lacking references that CAIN will have and therefore they too need to be supplemented for complete coverage.

Just recently, a retrospective search facility has been introduced and 100 retrospective searches of the full CAIN file of around half a million references have been carried out to date. From experience gained thus far, it appears that retrospective searches of CAIN are likely to average at about 25 per month, indicating substantial user interest in this search application.

Users of the ARS service are expected to formulate their own profiles (or retrospective queries) and initiate all revisions. It is generally accepted that this user self-reliance is likely to result in some weak profiles and without the presence of an interface between the user and the system it will not always be possible to identify the changes that are needed. The ARS search system is a modified version of the University of Georgia software, which is capable of searching a wide range of different data bases.

3.5.2 Quantitative studies of CAIN - No formal monitoring of the performance of any of the tape services or user reactions to them has been attempted.

3.5.3 User assessments of CAIN's strengths and faults - CAIN is thought to be particularly strong in its coverage of the foreign literature but the overall standard of coverage is considered to be poor. ARS expressed the view that the National Agricultural Library should have a more clearly defined policy on coverage to ensure that all items of major importance

are being included in CAIN. They are also critical of what they regard as an excessive time lag between the publication of some articles and their appearance in CAIN, and consider that improvements are needed in this area as well. Furthermore, ARS are critical of the need to rely on titles for searching and judging the relevance of hits and suggest that more subject terms and category codes should be assigned to items to give a better coverage of the important subjects contained in each document. Finally, based on their search experience with CAIN, they have identified various inconsistencies in the tapes such as the lack of English translations for some foreign language titles and the presence of duplicate entries for the same article, and consider that these faults could adversely affect search performance. However, it must be borne in mind that none of the views expressed above is supported by any quantitative evidence of CAIN's failings in specific areas but on subjective judgments based on experience gained in operating an SDI service.

3.5.4 Overall user reactions - The Agricultural Research Service are not entirely satisfied with CAIN but still continue to run more CAIN profiles than any other organisation covered in this study. Although CAIN is considered to be weak in several areas, they do recognise CAIN to be a necessary component of their information service because of its broad coverage of the agricultural literature. However, the extensive use of CAIN by USDA scientists is claimed to be due not to strong user satisfaction with CAIN, but to the realisation that preferred services such as Biological Abstracts and CAC need to be supplemented by CAIN for complete coverage. One reason which may partially explain user preference for Biological Abstracts and CAC is that the coverage requirements of USDA scientists are thought to be more heavily slanted towards the pure research and agriculturally related subjects than is normal at most organisations.

No data has been collected on CAIN's search performance or user reaction to it, but the general impression is that 75% of the total number of CAIN users find the service helpful and of this group, 25% are very satisfied with their results. In general, users react more favourably to the CAIN retrospective search facility than to the CAIN SDI service.

3.6 USER - Sentral for Forsoksmetodikk og Databehandling (FDB-sentralen), Vollebekk, Norway.

3.6.1 CAIN search experience - FDB-sentralen, a computer centre, have been operating a CAIN SDI service since October, 1971, in co-operation with the Agricultural College Library in Norway, for research scientists based at the College. They also run SDI services using Food Science and Technology Abstracts and World Textile Abstracts but do not search either tape in combination with CAIN.

When the CAIN service became operational in 1971, there were 120 CAIN profiles. A charging policy was introduced in January 1972, and this had the effect of reducing the number of profiles to around 70. During 1973, the number of profiles has dropped even further and at present, only 39 CAIN profiles are being processed by the Centre.

Users are required to construct their own profiles and initiate all profile revisions. FDB-sentralen recognise the advantage in having an information specialist trained in the vocabulary conventions of a data base to interact with users during the formulation of profiles, but have not as yet allocated staff for this purpose. Their CAIN search software appears to be more restrictive than most other user-developed search systems in that it places fairly severe limits on the size and complexity of profiles that can be processed.

3.6.2 Quantitative studies of CAIN - The following three studies have been completed by the Norwegian Computer Centre:

(i) Relevance studies - In 1971, an evaluation of the relevance of the search results produced for 69 CAIN profiles was carried out. These results indicate that even when items classified as being of secondary interest are taken into account, the average percentage relevance of the hits received by users was only 30%.

(ii) Questionnaire survey - Because initial reaction to CAIN was poor, a questionnaire survey was conducted to assess the extent of user dissatisfaction and reasons for CAIN's limited use by scientists at the College. The findings from this survey revealed that: (a) 80% of users contacted who were running CAIN profiles were satisfied with their search results; (b) users had stopped using CAIN for a variety of reasons with two of the principal reasons being inadequate relevance of the search results and poor coverage of the user's subject area, and (c) around 50% of the scientists who had never tried the CAIN service gave as their reason that they had not been aware that the service was available.

(iii) Coverage studies - FDB-sentralen have analysed the coverage provided by a 12-month portion of CAIN of articles selected from Scandinavian and Finnish agricultural journals. Their findings indicate that some scientific journals fail to get complete coverage (implying that articles of major importance may have been overlooked) and that a substantial portion of the total articles covered originate from non-scientific, advisory journals which are not likely to be of interest to their users.

3.6.3 User assessments of CAIN's strengths and faults - FDB-sentralen consider that although CAIN may well contain some publications which are difficult to locate from other sources, the general standard of coverage is poor. Their study of Scandinavian agricultural literature and user experience in searching CAIN has led them to the conclusion that CAIN contains too many popular articles, unlikely to be of interest to agricultural scientists and that there are cases when articles of major scientific importance fail to appear in the CAIN file. They are very critical of the need to rely on titles for both searching and when assessing the relevance of hits, but have not collected any quantitative evidence of CAIN's failings in either area. FDB-sentralen consider that the search performance of CAIN could be significantly improved by the addition of subject descriptors, to supplement titles when they fail to record all the important subjects covered by a document.

Finally, their investigation of the Scandinavian publications covered by CAIN has highlighted a number of inconsistencies in the conventions adopted for recording journal titles. Since journal titles are frequently used as search elements, FDB-sentralen are obviously concerned that such inconsistencies in format would result in failure to correctly match some of these titles during searches.

3.6.4 Overall user reactions - On the whole, FDB-sentralen users are dissatisfied with CAIN. The general impression is that CAIN provides inadequate coverage of the agricultural literature and its performance as measured by the relevance of search results is poor. However, there are no plans to discontinue the service.

(b) Information collected for five other users, extracted from reports and correspondence.

3.7 Alberta Information Retrieval Association, Edmonton, Canada - have had two years experience in running retrospective searches of CAIN and presently handle between 100 and 150 searches per year. Users are required to pay for searches but these charges have only discouraged users with a casual or border-line interest in the service. User reaction to CAIN is generally good.

3.8 University of California, Campus Computing Network, Los Angeles, California - have been running SDI searches of CAIN since January 1972, and as of December 1972 were searching over 100 profiles. Based on quantitative studies of relevance, they report an average relevance of 70% for their CAIN search results. A cost/benefit study carried out in 1972 showed that the majority of users found the CAIN service to be of material value and that benefits in terms of time saved by researchers far outweighed the overall costs of providing the service. User reaction to CAIN is generally favourable.

3.9 National Research Council of Canada (NRC), Ottawa, Canada - have been running experimental SDI searches of CAIN to familiarise the NRC staff with the structure and content of the CAIN file and are now preparing to make CAIN available across Canada through their CAN/SDI service. Since CAIN is not yet operational, it has not been possible to report on user reactions. However, NRC are convinced of the need to provide access to CAIN in order to obtain the necessary coverage of the agricultural literature.

3.10 The CAIN/CAB Project Group - at Nottingham, are studying the subject coverage provided by CAIN compared with two other data bases:

i Commonwealth Agricultural Bureaux (CAB) printed journals, and ii the BIOSIS tapes.

Their results show that the coverage of the CAB journals and the CAIN tapes overlaps by about 40% and that items generally appear in CAIN one to two months in advance of their publication in CAB. As expected, their work with BIOSIS indicates a small degree of overlap with the agricultural data bases but as yet no exact figures are available. Although this work is of undoubted interest, it is not of direct relevance to a study of user search experience with CAIN. Before any conclusions regarding search performance can be drawn from this work, it will be necessary to first run parallel searches of CAIN, CAB and BIOSIS and then to evaluate the search results in the light of the known subject overlap existing between them.

3.11 Chemie Information und Dokumentation (CID Berlin) - have analysed the content and format of a sample of records extracted from one issue of the CAIN tapes. CID Berlin are not regular subscribers to the CAIN service and studied it on a "one time only" basis. No information is given of any practical experience gained in using the CAIN service. As a result, their comments are of a theoretical nature only and are based on such a limited study as to be of little relevance to any evaluation of CAIN.

Other subscribers

The name of one remaining subscriber to the CAIN service for whom no information has been obtained is given in Section B.12.

Section 4. CONCLUSIONS DRAWN FROM THE FINDINGS

4.1 Interpretation of the main findings

When commenting on user reactions to CAIN, Information Centres tend to base their opinions on one or more of the following factors:

(i) Results obtained from quantitative studies - where the Centre has analysed such aspects as: (a) search performance, (b) coverage of the data base, and (c) benefits being derived from the service.

(ii) Statistical evidence of use - where growth or decline in the number of queries submitted can be indicative of user reactions to the service.

(iii) Subjective assessments and general impressions which are often not specific to CAIN but instead may relate to the overall computer-based service, of which CAIN is only one component.

In considering the various reactions to CAIN, it is necessary to keep in mind factors which may have influenced users either in their interpretation of statistical evidence or in the formation of their subjective judgments. As an example, poor profile construction could well result in poor search performance and subsequent user dissatisfaction quite independent of the file being searched. In the case of CAIN, there appears to be a degree of correlation between user satisfaction and the availability of an information specialist to act as an interface between the service and the user during the process of constructing profiles. When such an interface is provided, one can expect that the profile will reflect not only the requirements of the user, but the indexing and editorial policies adopted in the file being searched. As a consequence, one has to question whether the user dissatisfaction expressed by Norway and the US Department of Agriculture where users take responsibility for profile construction, is not at least partly due to a restrictive or weak profile rather than being a fundamental criticism of CAIN.

If the search software severely restricts the user as to the size and complexity of his question, then one can assume that the search performance will be reduced in some cases. For example, restrictive software may well be an additional factor which adversely affected the performance figures obtained for CAIN at the Norwegian Centre.

While extensive use of a data base can often be indicative of strong user satisfaction, there are factors which can alter this generalisation. For example, if users do not have to pay for searches it can be presumed that there will be some for whom the service is only of marginal benefit and in this situation a large number of profiles may not in itself reflect favourable user reactions. When alternative tape services are available, users will be less dependent on CAIN and therefore it is likely that fewer searches will be submitted than would be the case where CAIN is searched as the only data base. Finally, one must make a distinction between user need and user satisfaction when attempting to interpret statistical evidence of extensive use of a data base. For example, the US Department of Agriculture run more CAIN profiles than any other organisation covered in this

study, but they claim they do so, not because of strong user satisfaction with CAIN, but because of their need to obtain comprehensive coverage of the agricultural literature through CAIN.

A further factor to be considered is whether the user satisfaction that is expressed by a Centre reflects experience with a single data base or with a wide range of different tape services. In the latter case, user satisfaction is more significant since it is based on a relative assessment of CAIN versus other tape services.

4.2 Summary of main conclusions

In drawing conclusions regarding user reactions to CAIN, it is necessary to bear the above factors in mind.

While details of all known CAIN users have been presented in the Summary (Section 3) and Appendix B, it is clear that the results reported from CID Berlin and the National Research Council of Canada are not relevant to a study of user reactions since they do not reflect any search experience with CAIN. As a consequence, they will not be considered further in this report. Although the work being carried out by the CAIN/CAB Project Group is not directed at user search experience, their findings are relevant to this study since they deal with the important area of coverage.

In the remainder of this section, four specific aspects of user experience with CAIN will be presented. For each of these, there will be a general discussion, followed by a set of detailed conclusions.

4.2.1 Coverage

4.2.1.1 Discussion

Subscribers generally formulate their views on CAIN's coverage, not from any detailed knowledge of the articles that are contained in the file itself, but from an inspection of the references which they see in their search results. On the other hand, the work being carried out by the CAIN/CAB Project Group provides statistical evidence of CAIN's actual coverage in relation to other data bases. Their results seem to suggest that a substantial portion of the items contained in CAIN are not to be found in other services such as CAB and Biological Abstracts. However, they have also demonstrated that a large percentage of the items covered in an alternative agricultural service such as CAB fail to appear in CAIN.

Most Centres who run CAIN searches consider that CAIN provides good broad coverage of the agricultural literature. All agree that CAIN is particularly strong in it coverage of the foreign literature and the more obscure journals and non-conventional publications that are often difficult to locate from other sources. However, some Centres are critical of what is considered to be CAIN's excessive coverage of popular, advisory journals which are not likely to be of interest to research workers and some have expressed doubt as to whether all important scientific articles are being covered in CAIN. When the opportunity exists, Centres frequently search CAIN in combination with other agriculturally related data bases such as Biological Abstracts since they consider a multi data base approach is necessary for complete coverage of the agricultural literature.

Based on these two assessments of CAIN's coverage, the following conclusions can be drawn:

4.2.1.2 Conclusions

(i) CAIN provides good but not complete broad coverage of the agricultural literature containing articles of major importance as well as useful fringe material.

(ii) However, if comprehensive coverage is to be achieved, CAIN must be supplemented by alternative services including those which cover the agricultural literature within other scientific disciplines.

(iii) As a consequence of its breadth of coverage, CAIN contains a substantial amount of fringe material whose value may be assessed differently by subscribers, depending on the coverage needs of the individual user communities being served.

4.2.2 Accessibility

4.2.2.1 Discussion

The accessibility of items in a data base is affected by two principal factors:

(a) the amount and quality of the data given to express each item in the file, and
(b) the user's skill in formulating questions.

The first of these is controlled by the tape supplier, while the second is the responsibility of each individual subscriber. If the tape supplier fails to give a precise or comprehensive account of the important material contained within each document, then the user will often not be able to retrieve these items, regardless of how familiar he is with the vocabulary conventions of the data base.

There is evidence of considerable dissatisfaction with the methods used in expressing the subject content of items contained in CAIN because it is felt that there are insufficient enrichment terms and subject category codes. Furthermore, users find that in order to compensate for the lack of vocabulary control in the CAIN data base, they are often required to carry out complicated and tedious look-up and cross-checking procedures when selecting suitable search terms for profiles. A consequence of this is that in cases when users do not have the necessary skills or lack sufficient time to perform these tasks, CAIN's search performance may be adversely affected. It has been observed that when the responsibility for profile construction lies with the user, CAIN's performance was generally lower than was experienced by organisations where profiles were prepared by information specialists who combine a good knowledge of user requirements with an appreciation of how to access relevant items from a largley free-text file such as CAIN.

4.2.2.2 Conclusion

(i) Items in CAIN are often difficult to access and success in searching CAIN normally requires that considerable skill and effort be devoted to the formulation of the search question.

4.2.3 Suitability for SDI

4.2.3.1 Discussion

In this section, experience reported on the use of CAIN in providing an SDI service will be discussed under three main headings: Search performance, User surveys and Subjective assessments.

(1) Search performance - CAIN has been used most extensively in SDI systems but there are conflicting views on its search performance. The following quantitative evidence of CAIN's performance has been located from this study:

(a) Relevance - The Swedish Agricultural College (86 profiles/month) and the University of California (126 profiles/month) both report an average relevance of 70% for their search results while the Norwegian Agricultural College (39 profiles/month) have found that only a low percentage (30%) of hits supplied to their users were relevant to the questions posed. However, in the case of the Norwegian Centre, no assistance was provided to users in the preparation of their profiles and this will undoubtedly have had an adverse influence on the relevance figures obtained.

(b) Recall - The PUDOC (54 profiles/month) findings indicate that although CAIN may provide the best general coverage of the agricultural literature, it must be supplemented by alternative tape services that cover agriculture within other scientific disciplines if complete coverage of a user's subject field is to be achieved. It has also been demonstrated that for some profiles, these alternative services may outperform CAIN in terms of the number of relevant hits supplied.

(2) User surveys - Additionally, some users have formally monitored user reactions to their CAIN SDI service by conducting a questionnaire survey, and again there is conflicting evidence of user attitudes to CAIN. The following studies have been completed:

(a) Costs and benefits - Both the University of Florida (220 profiles/month) and the University of California results show that:

(i) the majority of their CAIN users consider the service to be of material value (e.g. time saved, duplication avoided), and

(ii) the reported benefits in terms of time saved by researchers far outweigh the overall costs of providing the service.

(b) General - The Norwegian survey highlighted a number of reasons for user dissatisfaction with CAIN and for the decline in the number of profiles being submitted to the Centre. However, the lack of trained staff to assist in profile construction will have had a bearing on the user reactions reported by the Norwegian Centre.

(3) Subjective assessments - Two other Centres make extensive use of CAIN for SDI purposes. University of Georgia (601 profiles/month) report no specific analysis of user reactions to CAIN, although the growth in the number of CAIN profiles

is comparable with the average growth in usage of the data bases searched at the Centre. The US Department of Agriculture (1041 profiles/month) are not entirely satisfied with CAIN but like the Georgia Centre have no quantitative evidence to support any claims of specific user reactions to CAIN. From their extensive use of CAIN, it can be presumed that CAIN is supplying references that are not being located from other sources and that it is therefore an important component of their information service. Their finding that it is necessary to search a combination of different tape services to achieve complete coverage of the agricultural literature is consistent with the PUDOC experience. It must also be remembered that the coverage needs of the main user population at the US Department of Agriculture are slanted more towards the agriculturally related fields than is normal at other Centres and as in Norway, the user receives no assistance in constructing his profile.

4.2.3.2 Conclusions

(i) CAIN can be used satisfactorily as an economically justifiable SDI service.

(ii) For general coverage of the current agricultural literature, CAIN appears to be the best available machine-readable data base, but for complete coverage, CAIN must be supplemented by alternative services.

(iii) In cases where user needs are slanted towards the pure research or agriculturally related subject fields, alternative services such as Biological Abstracts, Chemical Abstracts Condensates and MEDLARS are likely to provide more complete coverage. However, there is evidence of extensive and expanding use of CAIN for SDI purposes, even in cases where alternative tape services are outperforming CAIN, which suggests that CAIN is required if comprehensive coverage of the agricultural literature is to be achieved.

(iv) CAIN is capable of providing search results with a high percentage relevance for well-defined user profiles.

(v) User satisfaction with a CAIN-based SDI service is high when professional assistance is available for profile construction.

4.2.4 Suitability for retrospective searches

4.2.4.1 Discussion

The only reported experience with retrospective searching is that of University of Georgia (400 searches/year), Alberta (100-150 searches /year) and US Department of Agriculture (expect to run over 1000 searches/year). No formal monitoring of search performance has been attempted by any of these users and there have been no quantitative studies of user reactions. Alberta and US Department of Agriculture have both expressed favourable comments on CAIN's performance in retrospective searches, while there is evidence of substantial growth in the number of retrospective searches being submitted at Georgia and US Department of Agriculture.

4.2.4.2 Conclusion

(i) Although there have been no quantitative studies of this aspect of CAIN, the largely favourable user comments together with evidence of substantial and expanding use appear to suggest generally satisfactory experience with CAIN for retrospective searches.

4.3 Implications for AGRIS

It would be presumptuous to attempt to draw any detailed implications for AGRIS at present, since the author of this report is not in possession of sufficient information pertaining to the AGRIS project. However, it is possible to make a number of general observations on the basis of the conclusions arrived at from this study:

(i) Since agricultural science is an extremely broad discipline, it is unlikely that it can be served exclusively by a single data base. It can be anticipated that some users will find it necessary to supplement an agricultural data base with alternative tape services which cover agriculture within other scientific disciplines to achieve complete coverage of their field of interest.

(ii) The breadth of the agricultural subject area requires a data base covering an extremely wide range of publications. As a consequence of this, users may often encounter references in their search results which they consider to be of little value and may express annoyance at their inclusion in the file. It should, however, be borne in mind that this result is an inevitable consequence of attempting to achieve broad coverage and there is evidence from this study that users are prepared to accept this situation in the search for more complete coverage of their field of interest.

(iii) This study was concerned with a data base which provides a combination of titles and broad subject category codes as the means of representing the subject content of documents. The category codes enable the user to locate a group of documents which fall into his broad field of interest, while the free-text or title portions enable him to select articles from within this group which are likely to be of high relevance. It is apparent from this study that the category codes and free-text fields are complementary features of the data base and both have an important part to play in providing adequate recall and relevance. However, it should be borne in mind that the provision of sufficient enrichment terms (which need not be controlled) to supplement titles and an adequate number of category codes are essential to the ease of use and satisfactory performance of such data bases.

(iv) A tape service which is based on free-text and broad subject classifications can best be improved by introducing changes that reflect the inter-related role that each is to play during searches. It is better to combine adequate enrichments with a reasonably well-defined set of categories than to provide an extremely good classification system and no enrichments or numerous enrichments and no broad categories.

(v) It is possible that a data base of the type considered in this study could be substantially improved by the provision of very detailed classifications or controlled subject descriptors to supplement the free-text and broad classifications. However, there is virtually no evidence from this study to suggest that the lack of such subject control represents an insurmountable obstacle to the effective use of the data base. The findings appear to indicate that it is possible to obtain satisfactory results with a predominantly

free-text data base, given that users have sufficient skill and time to devote to the formulation of search questions.

(vi) There is strong evidence from this study that titles in the agricultural subject field generally give adequate information for the purpose of assessing the relevance of the search results.

(vii) The experience reported in this study indicates that a free-text data base, whose primary purpose is to provide an SDI service, can also be used satisfactorily for retrospective searches, given that suitable software is available.

APPENDIX A - DESCRIPTION OF THE B OF A SERVICE

A.1 SUPPLIER: The Bibliography of Agriculture (B of A) publications and tapes have been produced by CCM Information Corporation since 1970. The organisation name has recently been changed to: Macmillan Information, a Division of Macmillan Publishing Co.Inc., 866 Third Ave., New York, N.Y.10022.

A.2 DESCRIPTION OF DATA BASE: The printed and magnetic tape versions of the B of A are derived from the CAIN data base and each new edition contains articles received during the preceding month by NAL. A detailed study of the principal differences between CAIN and B of A and their effect on search performance has been completed by PUDOC. These findings are reported in Section B.3.6 of this report.

The B of A computer-based service consists of monthly tapes, each containing about 10,000 documents and software for running in-house searches. In addition to the various bibliographic fields that are extracted from the CAIN records, an additional field containing subject terms is generated for each item. These same subject terms appear in the printed subject indexes of the Bibliography of Agriculture and are automatically generated from document titles. A machine-readable dictionary is used as the basis for keyword selection as well as to control synonyms and word variants. The index term generation software has been adapted from a system previously developed for the PANDEX service (also supplied by Macmillan Information). The results of the PUDOC study indicate that the inclusion of controlled "title-derived" subject terms in B of A does not significantly affect search performance over that observed for CAIN.

The B of A tapes are supplied at a cost of $1000.00 per year in either PANDEX or MARC II format.

A.3 SEARCH SOFTWARE: A total of four SDI search systems are currently offered to customers of the Macmillan Information tape services. The various packages search either PANDEX or MARC II formatted tapes using either a boolean or weighted term search logic and all but one are designed for the IBM 360 series. Macmillan Information supply a search system, written in COBOL which can be used by subscribers with non-IBM equipment.

A.4 SUBSCRIBERS TO THE SERVICE: Four subscriber names have been obtained from Macmillan Information:

(i) Centre for Agricultural Publishing and Documentation
PUDOC,
Wageningen, The Netherlands.

PUDOC's experience with the B of A tape service is outlined in Section B.3 of this report.

(ii) Ohio State University Library,
Mechanized Information Center,
1827 Neil Avenue,
Columbus, Ohio 43210

The Ohio State University currently run SDI searches of B of A for 87 scientists, most of whom are based at the College of Agriculture. The B of A service was introduced in April, 1973 and new users are being added at the rate of 20 per month at the present time. No quantitative

information has yet been obtained on search performance or user reactions, but the overall impression is that B of A is on a par with the other tape services which are being searched.

(iii) Kinokuniya Book-Store Co. Ltd.,
826 Tsunohazu 1 CHOME,
Shinjuku-Ku,
Tokyo, 160-91 Japan.

No information is available.

(iv) Ames Laboratory,
Computer Building,
Iowa State University,
Ames, Iowa 50010.

No information is available.

APPENDIX B - DETAILS OF USER SEARCH EXPERIENCE WITH CAIN

B.1 USER: LANTBRUKSHOGSKOLAN, ULTUNABIBLIOTEKET, 750 07 Uppsala 7, Sweden.

B.1.1 SOURCE OF DATA: (i) Visit with B. Rufelt at the Agricultural College in Uppsala, August, 1973. (ii) Two letters from B.Rufelt, May and September, 1973. (iii) Report entitled "CORSAIR, A computer program system for information retrieval" by Sten-Sture Tersmeden, FOA 2 rapport, C 2553-M7, August 1972.

B.1.2 CAIN SEARCH EXPERIENCE: The Library of the Agricultural College in Uppsala has been providing an SDI service based exclusively on CAIN since January, 1971 for scientists working at the College as well as other Institutes, Universities and commercial organisations throughout Sweden. Low acquisition cost and good broad coverage of the agricultural literature were given as the principal reasons for the initial selection of CAIN as the basis for an SDI service. The service is run in co-operation with the FOA Index (a section of the Research Institute of Swedish National Defence) who act as a computer centre, supplying the necessary software and machine facilities for execution of the various search runs. FOA Index have wide experience in searching a number of commercial tapes in addition to CAIN.

The CAIN SDI service was offered free of charge to users for the first 24 months of operation but since January 1973, it has been running on a cost recovery basis. In December 1972, when the service was still free, the Library was handling 119 user profiles while in the next month, when the charging policy was introduced, the number dropped to 72 profiles. In August 1973, a total of 86 SDI profiles representing 70 users were being processed, and each user was charged at a rate of £20 to £35 per profile per year, depending on the number of hits obtained during the year for his profile. Since 1971, only about five retrospective searches of CAIN have been carried out and in each case, only a six month portion of the cumulative file was examined. The lack of user interest in computer-based retrospective searches and the high machine running costs likely to be incurred when processing large files were the main reasons given for the limited use of CAIN in this area.

Since first implementing the SDI service nearly three years ago, the Library has continued to collect feed-back from users as to the relevance of their search results. An analysis of user relevance assessments for the period of operation from January 1971 (when there were ten profiles) to December 1972 (when there were 72 profiles) has been carried out, and the following statistics are reported:

(see next page)

Interest ratings from users	1971 Results		1972 Results	
	Refs.	Per cent	Refs.	Per cent
4. Of immediate interest	1350	20,5	2024	18,6
3. Relevant	2455	37,4	4522	41,6
2. Of secondary interest	802	12,2	1091	10,1
1. Not relevant	1759	26,8	2805	25,8
0. Cannot determine interest because the title does not provide enough detail	203	3,1	422	3,9
	6569	100,0	10864	100,0

As can be seen, the 1971 figures indicate that on average, 58% of the hits obtained by users during the year were of prime importance, while an additional 12% were of secondary interest. Similar trends appear in 1972, when 60% of the hits were classified as being important and 10% were considered to be of secondary value. Including items classified as being of secondary interest, an average relevance of 70% has been demonstrated for search results obtained during this two year period. On average, a user receives about 150 answers to a profile throughout a given year and approximately 100 of these (i.e. 70%) are of major or secondary interest to him. It is noteworthy that titles proved to be inadequate for the purpose of assessing relevance for only about 4% of the total number of hits obtained. These results conflict with the views expressed by some of the other organisations running CAIN searches, who have postulated that a weakness of a title-based service such as CAIN is that the user is often severely limited by the lack of detail given in titles when assessing the relevance of the search results.

Some preliminary relevance statistics have been obtained for 45 of the 81 profiles which were being processed during the first three months of 1973, and the findings are tabulated below.

Profile No.	No. of hits	No. of relevant hits (excl. those of secondary interest)	%age of relevant hits
1	156	76	48
2	8	8	100
3	23	22	95
4	120	70	58
5	34	16	47
6	61	37	40
7	27	22	81
8	32	28	87
9	16	7	43
10	3	3	100
11	111	73	65
12	24	7	29
13	11	6	54
14	55	3	60
15	77	16	20
16	49	36	73

(continued)

Profile No.	No. of hits	No. of relevant hits (excl. those of secondary interest)	%age of relevant hits
(continued)			
17	142	34	23
18	23	55	21
19	4	4	100
20	99	32	32
21	26	25	96
22	30	4	13
23	20	18	90
24	210	174	82
25	63	63	100
26	63	61	96
27	27	4	14
28	69	40	57
29	7	3	42
30	2	2	100
31	2	2	100
32	110	24	21
33	47	31	65
34	3	3	100
35	65	29	44
36	18	3	16
37	80	72	90
38	60	57	95
39	38	36	94
40	6	1	16
41	12	3	25
42	25	14	56
43	59	20	33
44	3	2	66
45	27	18	66

From these figures, it is possible to categorise the profiles according to the relevance rating achieved for the search results during the three month period. The profile groups are listed below.

% of hits obtained for the profile that were relevant	No. of profiles corresponding to the relevance rating
0-10	0
10-20	4
20-30	6
30-40	2
40-50	6
50-60	4
60-70	5
70-80	1
80-90	3
90-100	7
100	7
	45

These statistics indicate that about half of the total number of profiles analysed were giving search results in which 60% or more of the hits obtained were relevant to the question posed

The search output contained relatively few relevant hits for a quarter of the profiles studied, while for just under one third of the total number of profiles a relevance of between 80 to 100% was observed for the search results.

B.1.3 PROFILE CONSTRUCTION: All user profiles are constructed by the Library staff who first check the suitability of CAIN for any new queries that are submitted by conducting a trial search using the Bibliography of Agriculture (printed version). When a question is considered to be within the subject scope of CAIN, it is run free of charge to the user for at least three months, while the initial profile formulation stage is completed. During this time, the search results are compared with references obtained from duplicate searches of Bibliography of Agriculture and other printed services in an attempt to highlight areas in which the profile should be improved to provide more precise or comprehensive coverage of the user's field of interest. On average, each profile contains about 20 search terms where a term can be entered in full or as a word root.

B.1.4 USER SEARCH SYSTEM: The CORSAIR III (Computer Oriented Reference System for Automatic Information Retrieval) system developed by FOA Index is a general purpose IR package with a wide range of applications, one of which is the searching of commercial tapes such as CAIN. The software runs on an IBM 360/75 at the FOA Index computer centre and programs are written in Fortran and Assembler. The CAIN tape records are converted to a MARC II type of format before being searched. All data elements in each record are searchable but in practise, search terms generally relate to the title, category code and subject term fields and in a few cases other data types such as author, journal title and language codes are included in the search statement. The search logic is flexible, allowing for an unlimited degree of nesting in the Boolean statement and there are virtually no practical restrictions on the number of terms allowed per profile. The software incorporates various optimisation features which permit efficient serial scanning of the textual data contained in each CAIN file, resulting in fairly low running costs. The machine running cost associated with searching a batch of 81 profiles comprising 1449 search terms against one monthly CAIN tape was reported to be about £70, corresponding to a cost of about 85 pence per profile per run. Out of 36 tapes processed to date, FOA Index have received only one faulty tape which had to be returned to NAL.

B.1.5 USER ASSESSMENT OF CAIN: On the whole, user reaction to CAIN is very favourable. The results of detailed relevance studies carried out over a period of nearly three years indicate that the search performance measured in terms of the percentage relevance of search results obtained is relatively high. As reported previously both the average figures taken over a two year period and relevance data for individual profiles studied in 1973 show that a high proportion of the total number of hits supplied to users are relevant to the questions posed. In general, users rely on CAIN to a very large extent for their current awareness needs, but most feel the need to supplement CAIN by other services to get complete coverage of their subject fields.

B.2 USER: UNIVERSITY OF FLORIDA, Institute of Food and Agricultural Sciences (IFAS), Hume Library, Gainesville, Florida.

B.2.1 SOURCE OF DATA: (i) Visit with A.C. Strickland at University of Florida, September 1973. (ii) Letters from A.C. Strickland, August 1973, and G.T. Kovalik, September 1973. (iii) Report entitled "CAIN SDI service - a user survey" by G.T. Kovalik, June 1973.

B.2.2 CAIN SEARCH EXPERIENCE: Hume Library is now in its fourth year of supplying monthly SDI printouts from CAIN for research workers employed by IFAS, a division of the University of Florida. In addition to the College of Agriculture in Gainesville, IFAS incorporates 22 experimental stations scattered throughout Florida with many of these stations providing only very limited library facilities. The CAIN search service is run in co-operati with the University's Computer Department who have developed software specifically for this SDI application. Good broad coverage of the agricultural literature and low acquisition costs were reported to be the principal factors that led to the choice of CAIN over other commercially available tapes.

The Library is currently processing 220 CAIN profiles on behalf of about 200 scientists and no charge is levied for searches. Roughly one half of the scientists within IFAS who could be regarded as potential CAIN users are in fact utilising the service. The number of profiles being searched has more or less reached a constant level by now, after rising sharply during the initial period of operation, but the trend has always been towards increased usage of the service. The University of Florida have not had any experience in running retrospective searches of CAIN but are planning to take advantage of the Lockheed on-line service in the near future (see Section 2.5). It is felt that the on-line capability for running retrospective searches would be of considerable benefit to the various extension services throughout the State of Florida.

The Library have never at any time attempted to formally monitor the relevance of search results, but are of the general impression that about 70% of the total number of hits retrieved are relevant to the questions posed. However, it has been their experience that there is not always a clear relationship between the performance of the search as measured by relevance and user reactions to the service. They have found that a high percentage relevance for the search results does not in itself ensure that a user is deriving benefit from the service, and conversely, there are many examples of favourable user reaction in cases where the percentage of relevant items in the SDI printouts received each month is extremely low. Often the relevant hits that are obtained would have been very difficult to locate from other sources and this factor normally outweighs the inconvenience to the user of having to visually select relevant items from the total list of answers.

B.2.3 PROFILE CONSTRUCTION: All user profiles are prepared by the Library staff in co-operation with the scientists concerned. The Library use a wide range of printed services including the Bibliography of Agriculture to assist them in the initial construction of profiles and in cross-checking CAIN search results in an attempt to identify any faults in the profile. The library staff regularly visit the research scientists at the College and at the different experimental stations to assess user reactions to the service, and if the reaction is unfavourable, take the opportunity to discuss changes that might be made to the user's profile.

On average, a profile contains about 70 terms. This larger than normal profile size is mainly due to the lack of a word truncation facility which therefore requires that each word variant be expressed in full as a separate search term.

B.2.4 USER SEARCH SYSTEM: The University of Florida CAIN search package is written in COBOL and is run on an IBM 370/160. The following fields in each CAIN record are searchable: titles, Subject descriptors, category codes, authors and journal titles, but in practise, only titles and subject descriptors are actually being searched. The category codes are ignored as possible search terms, mainly because there is little confidence that the codes are being assigned consistently by NAL, while the need for searching the other data types such as authors has not yet arisen. At present, only the Boolean operators AND and OR may be used, although there are plans to incorporate "BUT NOT" logic in the near future. Other than this limitation, the search logic is flexible allowing an unlimited degree of nesting in the Boolean statement, and there are virtaully no practical restrictions on the number of terms allowed per profile. The machine running cost per year is currently ₿3400.00 (for 220 profiles) and the cost per profile per monthly run is therefore ₿(3400/220 x 12) or about 50 pence. Over a period of four years, the University of Florida have received a total of three damaged tapes from NAL which had to be returned.

B.2.5 USER ASSESSMENT OF CAIN: On the whole, IFAS users are very satisfied with their CAIN search results. Users with access to library facilities normally check other sources to supplement these results while scientists at remote experimental stations throughout Florida rely solely on CAIN for their current awareness needs.

In 1973, the Library undertook a questionnaire survey of user reactions to CAIN in order to obtain some quantitative evidence of benefits that were being derived from the service. A total of 184 scientists were contacted and there was a 73% response. The principal results of the survey are summarised below.

Questions	Answers	
(a) In general, are the CAIN printouts of material value (i.e. time or effort saved, duplication avoided) to your research needs?		
(i) Yes	108	(80%)
(ii) No	27	(20%)
(b) If this monthly printout continues for the next several years, what sort of impact might it have on your literature surveillance habits?		
* (i) I would save ___ hours of search time	83	(59%)
(ii) Not much effect	27	(20%)
(iii) Negative effect	3	(2%)
(iv) Other	26	(19%)

* 70 filled in hours saved (non-numerical answers were discounted)

(see next page)

Hours saved	No. of responses	Man Hrs./week
1	16	16
2	30	60
3	8	24
4	8	32
5	5	25
6	1	6
7	1	7
8	-	-
9	-	-
10	1	10
		180

(c) Would you be willing to pay for the CAIN service?

(i) Yes	88	(51%)
(ii) No, library should provide the service free of charge	66	(49%)
(iii) No, explain	17	

(d) Do you think a different data base would suit your needs better?

* (i) Yes	24	(25%)
(ii) No	72	(75%)

* If yes, which of the following?

Biological Abstracts	24	
Chemical Abstracts	5	
Engineering Index	4	
Other: Agricultural Index	1	
Bibl of Reproduction	1	
Helminthological Abstracts	1	
Food Sci. & Technol Abstracts	1	
Abstracts of Entomology	2	
Rev. Applied Mycology	1	
Rev. Plant Pathology	1	

Finally, users were asked to comment generally on their reactions to the CAIN service. A selection of the replies received appears below.

(i) CAIN gives very good coverage of the foreign literature.

(ii) No time saved, but CAIN helps me find articles that I might otherwise miss.

(iii) As library facilities are not available here at Ft.Lauderdale, this service is of immense value to me. This is a fine service for the branch experimental stations.

(iv) The CAIN service is designed to fit individual needs and as such is superior to generalised abstracting services.

(v) A quick way to learn of recently published literature.

(vi) It provides references on which I can back-track when making a detailed literature search on a specific subject.

(vii) It is extremely valuable as a means of keeping informed of current literature which is pertinent to my field.

(viii) I don't think CAIN is fully comprehensive - but neither is any abstracting or indexing service.

(ix) It turns up too many foreign items with no translation available.

(x) I get very few relevant answers so it is of little value to me.

The results from this study indicate that the majority of users (80%) find the CAIN service to be of material value to them particularly in terms of hour of research time saved. The service appears to be of special importance to scientists located at isolated research centres where CAIN may be the only major link with the literature being published in their subject fields. Judging from the response to question (c), the effect of introducing a charging policy for searches would be to reduce the number of profiles by about 50%. This is consistent with the experience of both the Swedish and Norwegian Colleges (see Sections B.1.2 and B.6.2) who presently run their CAIN services on a cost recovery basis.

Based on the statistics collected for question (b), a cost/benefit analysis was carried out. The results from this question indicate that the total time saved by 70 IFAS scientists as a result of the CAIN service was 180 hours per week, representing a total of 9360 man hours per year. This figure can be translated into a cost saving of $65,520, assuming a value of $7.00 per hour for the researcher's time. The total cost per year of operating the system including administrative and service overheads, software development, and staff effort amounts to $17,822. Thus one can conclude that the benefits derived from CAIN in terms of research time saved, far outweigh the cost of providing the service. However, it can be argued that the time saved as a result of using CAIN is not necessarily going to be applied productively to research projects, and therefore the cost savings cannot be accurately defined without a more detailed knowledge of the working habits of the users concerned.

B.3 USER: CENTRE FOR AGRICULTURAL PUBLISHING AND DOCUMENTATION (PUDOC), Wageningen, the Netherlands.

B.3.1 SOURCE OF DATA: (i) Visit with H.C. Molster at PUDOC in August 1973. (ii) Report entitled "Testing a scheme for machine-retrieval of information in agricultural science" by D.J.Maltha, PUDOC. (iii) Report entitled "Differences between the CAIN and Bibliography of Agriculture tapes" by H.C. Molster, August 1973.

B.3.2 CAIN SEARCH EXPERIENCE: The Centre for Agricultural Publishing and Documentation, better known as PUDOC, provide as one of their functions, a national information service for agricultural literature. In the past, it has relied mainly on externally prepared printed bibliographies supplemented by a number of internal, manually operated systems. Since 1972, PUDOC have been evaluating the usefulness of a computer-based information service in satisfying the current awareness needs of Dutch agricultural scientists. A number of commercial tape services including CAIN and the Bibliography of Agriculture (B of A) are being tested on an experimental basis with users reporting back on the relevance of hits received, to allow an assessment of the performance of the various tapes searched. The service is being offered on a trial basis only at present, and it has not been integrated with the different manual, current awareness systems in use within the PUDOC Information Department.

PUDOC began using CAIN in September 1972, and are currently running 54 SDI profiles against CAIN and different combinations of other tape services at no charge to users. Throughout the period of operation, data has been collected on the relevance

and recall of the search results obtained from the different tape services and those findings which are applicable to CAIN will now be reported.

The first comparative study of search performance to be completed by PUDOC used relevance as the only criterion for assessing the different tape services, and it was based on an analysis of nine SDI profiles. The main findings from this investigation are summarised below:

(a) Study of nine SDI profiles

Nine SDI profiles in the agricultural and veterinary science field were run against the CAIN and B of A tapes and also BA Biological Abstracts), BRI (Bio Research Index), CAC (Chemical Abstracts Condensates), FSTA (Food Science and Technology Abstracts) and MEDLARS (Medical Literature Analysis and Retrieval System). The nine profiles were run against the following combination of tapes in the period from March, 1972 to Jan, 1973:

Title of profile	Tapes searched
1. Phosphate requirement of livestock	*MEDLARS, CAIN, B of A, CAC, BA, BRI
2. Control of sex in plants	*CAIN, B of A, BA
3. Influence of fluorides on plants	*CAC, CAIN, B of A, BA, BRI
4. Coccidiostats	*MEDLARS, CAIN, B of A
5. Protein content of cassava	*FSTA, CAIN, B of A, BA
6. Lipoxidase in seeds	*CAIN, *B of A, BRI, BA, CAC
7. Preparation and use of protein	*B of A, CAIN, CAC, FSTA
8. Aromatic constituents of foodstuffs	*CAIN, B of A, CAC, FSTA
9. Pesticide analysis of foodstuffs	*MEDLARS, CAIN, B of A, FSTA, CAC

* = tape against which highest percentage of relevant hits were obtained.

The relevance of hits produced from running the nine SDI profiles against the various tapes was assessed and the findings are summarised below.

Tape	No. of profiles tested	No. of profiles for which relevant hits were obtained	No. of profiles where tape provided highest %age of relevant hits (compared with the other tapes)
CAIN	9	7	≠ 3
MEDLARS	3	3	3
B of A	9	7	≠ 2
CAC	6	6	1
FSTA	4	4	1
BA	5	4	0
BRI	3	3	0

≠ For one profile, CAIN and B of A shared the highest rank since all hits resulting from each tape were relevant.

Based on this relevance data, the following main conclusions can be drawn:

(i) CAIN and B of A provided relevant hits for most of the profiles tested.

(ii) For each of the nine profiles tested, one or more of the "non-agricultural" tape services produced relevant hits. A comparison of the CAC search results with the CAIN and B of A results for one of the profiles, indicates that some "CAC relevant hits" were not retrieved from the CAIN and B of A searches.

(iii) For profiles which fall into the food sciences subject area, both CAIN and B of A performed particularly well. It is noteworthy that for profile No.8 (aromatic constituents of foodstuffs), CAIN gave a higher percentage of relevant hits than did FSTA.

(iv) A comparison of hits obtained from searches of CAIN and B of A against the same profile appear to indicate that there are substantial differences in coverage between the two services. A further comparative study of these two tape services has recently been completed, and the main findings are reported below in section B.3.6.

(v) MEDLARS produced search output having a very high relevance for all of the profiles against which it was tested but the recall from these searches may have been low.

(vi) CAC and FSTA performed well for most of the profiles tested. BA and BRI provided few relevant hits and appear to be less well suited to the agricultural and veterinary science subject field.

The original investigation has been extended to cover a group of 54 SDI profiles and is now concerned with other measures of search performance in addition to relevance. Unfortunately, this second investigation has not yet been completed and it is only possible to give a general indication of progress made thus far.

(b) Study of 54 SDI profiles

The main aims of the current investigation are:

(i) to compare the results obtained from manual and computer-based current awareness searches using recall and timeliness as the basis for comparison, and

(ii) to compare the relevance, recall and timeliness of SDI results produced from the different tape services.

Users are required to report back on the relevance and timeliness of references received in their SDI printouts to allow an assessment of the relative performance of the different tapes searched. All profiles are searched against CAIN and B of A and some are searched against the other available tape services depending on the subject content of the profile.

Tape service	No. of profiles searched against the tape
CAIN and B of A	54
CAC	16
FSTA	15
MEDLARS	6
BA and BRI	3

As can be seen from the above table, CAIN and B of A are the principal components of the SDI service as far as usage is concerned.

Based on relevance data collected for 52 CAIN profiles, PUDOC report that 35 profiles are currently producing search results having a percentage relevance of 70% or higher. The percentage relevance of the search results obtained for the remaining 17 profiles is between 20 and 70%. It was stressed that relevance should not be used as the only measure of search performance since there are cases when a user may react favourably to a service even though the percentage of relevant items received from that service may be extremely low. The user may be very satisfied with the few relevant hits obtained if they are items which he would not otherwise be able to locate.

Only a very limited evaluation of recall has been attempted so far. No data is yet available on the relative recall provided by manual versus mechanised systems but some comparisons of the recall observed for different tape services have been carried out. The findings from two studies which include CAIN are outlined below:

Profile 1 - Potato products for consumption

Searches of the 1972 FSTA and CAIN tapes showed that:

(a) FSTA yielded 57 relevant hits not located from searches of CAIN.

(b) CAIN yielded 15 relevant hits not located from searches of FSTA.

(c) The same four relevant hits were retrieved by both data bases.

Searches of the 1973 tapes showed that:

(a) No relevant hits were obtained from FSTA that were not also located by CAIN.

(b) CAIN yielded four relevant hits not located by FSTA.

(c) The same two relevant hits were retrieved by both.

This study seems to indicate that both CAIN and FSTA are required for comprehensive coverage of the food science subject area since each provides relevant hits that the other misses.

Profile 2 - Influence of aerial fluorine on plants

Searches of the 1972 CAIN, BA and CAC tapes showed the following trend:

Data base	Relevant hits that were unique to the data base
CAIN	18
CAC	23
BA	23

In some cases, the same relevant hits were retrieved by more than one data base and these occurrences are summarised below:

Data bases	Relevant hits that were unique to the data base group
CAIN & CAC	2
CAIN & BA	4
CAC & BA	6
CAIN & CAC & BA	5

As with the previous study, these results emphasise the need for a multi data base approach if comprehensive coverage of the agricultural literature is to be achieved.

No information on user assessments of the relative timeliness of references provided by the different tape services is available at the present time.

PUDOC have not had any experience in running retrospective searches of CAIN or any of the other data bases.

B.3.3 PROFILE CONSTRUCTION: All profiles are constructed by an information specialist who acts as the interface between the user and the different tape services. When preparing a CAIN profile, he consults a wide range of printed services such as Bibliography of Agriculture in the hope that they will highlight appropriate search terms to be included in the profile. The results of early search runs are checked against alternative services in an attempt to identify any weaknesses in the profile which if improved, would result in more precise or comprehensive coverage of the user's field of interest. On average, each profile contains about 20 search terms where a term can be entered in full or as a word root.

B.3.4 USER SEARCH SYSTEM: PUDOC run CAIN, B of A and FSTA as in-house services in co-operation with IWIS-TNO (TNO Institute for Mathematics, Information Processing and Statistics), the Hague, who supply the necessary software and machine facilities for execution of the different search runs. To provide a comprehensive service for their users, PUDOC have made arrangements with two organisations to gain access on a service bureau basis to the remaining tapes (CAC, MEDLARS, BA and BRI). CAC searches are run in the Hague at the IWIS-TNO while the remaining tape searches are carried out at the Biomedical Documentation Centre, Karolinska Institutet, Stockholm. The description of software which follows applies only to the system developed for searching the CAIN, B of A and FSTA tapes which is known as FLIRT (Free Language Information Retrieval System).

FLIRT has been developed by IWIS-TNO and is based on a system previously designed for the INSPEC and CAC tapes. The software runs on a CDC-3200 and programs are written in Fortran. The CAIN tape records are converted to a MARC II format before being searched. The following data elements can be searched in each CAIN record: title, category codes, subject terms, authors, document type code and translation code. The search logic provides the normal AND, OR, BUTNOT Boolean operators as well as a weighted search facility. The software does not permit the use of parenthesis in the Boolean statement (i.e., no nested logic). For example, the system could not handle the question - A OR (B AND C) OR D. There are virtually no limitations on the number of search terms allowed per profile.

The machine cost associated with searching CAIN is estimated to be about £7 per profile per monthly run. Since beginning their CAIN SDI service in September, 1972, PUDOC have received two faulty tapes which had to be returned to NAL.

B.3.5 USER ASSESSMENT OF CAIN: PUDOC users are generally satisfied with the CAIN service and consider that it provides good broad coverage of the agricultural literature. Their first assessment of the different tape services, where relevance was used as the only measure of search performance, indicated that CAIN and B of A performed slightly better than the other tape services when viewed in the context of the total number of profiles tested. CAIN and B of A yielded relevant hits for 7 of the 9 profiles included in the study and for three profiles produced results which had a higher percentage relevance than those obtained from other tape services. However, in cases when a profile fell into the scope of "non-agricultural" data bases such as MEDLARS and CAC, these data bases often outperformed CAIN in terms of the percentage of relevant hits provided. Their subsequent studies of the relative recall obtained from CAIN and other tape services support the main finding of this earlier study which is that although CAIN is a necessary component of any machine-based agricultural service, it needs to be supplemented by other data bases which cover the agricultural literature within other scientific disciplines.

The search performance of CAIN, as measured by the percentage of relevant hits obtained for the existing 54 profiles, is relatively high. The present figures indicate that a high proportion of the hits received by a user are relevant to the question posed.

B.3.6 USER ASSESSMENT OF CAIN vs B OF A: PUDOC have recently completed a separate study of the main differences between CAIN and B of A (the contents of the latter being derived from the CAIN data base). The effect on search performance arising from the following differences between the two tape services was analysed:

(i) Coverage - B of A does not include the monographs and new serial titles that are contained in CAIN.

(ii) Foreign titles - B of A does not give foreign language titles (only the English translation) while CAIN does.

(iii) Descriptors - B of A includes descirptors (generated automatically from titles) and these are not contained in CAIN records.

An average of 27 profiles were run against both services for a period of seven months and the following search results were obtained:

(i) No. of hits retrieved from both	2485 (78%)
(ii) No. of hits from CAIN only - because of foreign language titles	68 (2%)
(iii) No. of hits from CAIN only - because of monographs and serials	336 (11%)
(iv) No. of hits from B of A only - because of descriptors	177 (6%)
(v) No. of hits unique to B of A or CAIN, which are unexplained due to software problems	100 (3%)

As it was known that B of A contains some duplicate entries in cases where more than one category code has been assigned, a check on the duplicate hits contained in the B of A search results was carried out. PUDOC found that 206 of the hits retrieved from B of A (corresponding to 1% of the total) were duplicates.

A further study of those hits which had been produced from B of A and not by CAIN because of the presence of descriptors in B of A indicated that a negligible number (0.5%) were relevant to the questions posed. PUDOC conclude from this that the presence of descriptors in B of A does not give B of A a significant advantage over CAIN in terms of search performance.

Based on these findings, PUDOC have drawn the overall conclusion that the inclusion of foreign titles and monographs in CAIN is of greater importance to them than the additional descriptors provided by B of A and as a consequence, CAIN is the preferred data base.

B.4. USER: UNIVERSITY OF GEORGIA, Office of Computing Activities, Athens, Georgia 30601

B.4.1 SOURCE OF DATA: (i) Visit with M.K. Park at University of Georgia, September 1973. (ii) Letter from M.K. Park, to Dr. U. Schutzsack (IFIS) Dec. 7th, 1972. (iii) Final report on NSF Grant GN-851, Dec. 1972.

B.4.2 CAIN SEARCH EXPERIENCE: The Information Services Division of the Office of Computing Activities have been offering computer-based information services to scientists within the 30 Institutions in the University System of Georgia, as well as other academic, government and commercial organisations for a number of years. The Centre is presently searching 20 commercial tape services for a total of 1661 users where users are defined as mailing addresses. Over 9000 SDI profiles are currently being searched and a total of 6624 retrospective searches have been run by the Centre from June 1972 to June 1973. The service is run free of charge to scientists within the University System and on a cost recovery basis to outside users.

The Georgia Centre started its use of CAIN in 1970 by purchasing both the 1969 CAIN tape (containing pesticide material from 1967-69) as well as the normal monthly service. They currently run 601 CAIN SDI profiles of which 596 correspond to scientists within the University System of Georgia, while the remaining five profiles are being run at a cost of about £50 per profile per year for users outside of the University System. Several advisory or extension services also make use of the CAIN search service but they represent a small percentage of the total user group. The increase in the number of profiles run against CAIN since the service was first introduced is summarised below:

Year	No. of CAIN profiles
June 1970 - June 1971	196
June 1971 - June 1972	395
June 1972 - August 1973	601

The growth in the number of retrospective searches run against CAIN can be observed in the following table:

Year	No. of CAIN Retro.Searches
June 1970 - June 1971	179
June 1971 - June 1972	224
June 1972 - June 1973	estimated to be around 400

In August, 1973, the Georgia Centre ran 37 retrospective queries against CAIN which if extended to a yearly rate would imply that the total number of retrospective searches in the following year is likely to be at least 500. CAIN retrospective searches are run at two-weekly intervals, and the batch size presently averages at around 20 queries.

As of August 1973, their cumulative CAIN file was found to contain 415,896 records. This figure has been derived from the following yearly statistics:

Year	Size of CAIN file
1969 (pesticide material only)	42667
1970	39918
1971	121782
1972	143485
1973	68044
	415896

In general, the number of searches being run against CAIN at present is near to the average for the other data bases. This trend is likely to continue since the number of new CAIN queries is increasing at a rate which is more or less in line with the average growth observed for all the data bases.

No attempt has been made to monitor the performance of any of the tape services and very little user feed-back has been communicated to the Centre regarding the relative advantages of different tapes.

B.4.3 PROFILE CONSTRUCTION: A single profile (or retrospective query statement) is prepared per question regardless of the number of data bases that are to be included in the search. In most cases, profiles are constructed by an information specialist who is familiar with the indexing and editorial policies followed by the different data base suppliers.

No specific information was available on the procedures adopted when formulating profiles which are to be run against CAIN. In general, the first step in preparing a profile involves the selection of any terms which are to be located in the free-text areas of each tape record. These terms are then supplemented by the appropriate classification codes or subject descriptors which apply to the data bases included in the search. The information analyst normally reviews the results produced from the preliminary SDI runs to assess the overall search performance achieved. If necessary, the profile is revised, but no monitoring of the search results produced from subsequent runs is attempted. The user is expected to recognise the need for further profile revisions and to seek the necessary guidance for implementing these changes. The Georgia system provides an on-line interactive data entry system for editing the user profiles, but the subsequent searches are carried out in a batch mode.

B.4.4 USER SEARCH SYSTEM: University of Georgia have developed a general purpose IR system (UGA Text Search System) for searching the multiple data bases included in their service. Each data

base is converted to a standard, generalised format, which then allows the file to be processed by the search system.

The system is presently operating at the Georgia Centre on an IBM 360/65 and programs are written in PL/1 and Assembler. All data elements which occur in a tape record are searchable. The software, which is used for both SDI and retrospective searches, provides the normal Boolean logical operators as well as a weighted search facility. There are virtually no restrictions on the number of terms permitted per profile or on the level of nesting in the Boolean statement. The search software incorporates a number of optimisation routines which have been designed to provide efficient processing of large bibliographic files. As an indication of search efficiency, the machine running costs associated with the last search of the entire retrospective CAIN file (of nearly half a million references) against a batch of about 18 queries was £125, corresponding to a cost of about £7 per question. A comparison of the machine times for searching the same number of documents in data bases such as Biological Abstracts (BA) and Chemical Abstracts Condensates (CAC) with the times observed for CAIN indicate that CAIN searches are 3 to 5 times more efficient. This time difference is attributed to the fact that BA and CAC records contain more information per document, resulting in longer length records (roughly twice the size of CAIN records) that need to be scanned during searches.

B.4.5 USER ASSESSMENT OF CAIN: University of Georgia take a fairly neutral view regarding all the tape services which they search and were reluctant to comment on their reactions to CAIN.

In order to assess the impact of the multi data base service on their users, a questionnaire survey was conducted in 1972. The findings indicated that over 97% of the users who responded felt that the service contributed significantly to their professional activities with the major benefits being time savings and broadened subject coverage. The survey was concerned only with the overall service and did not set out to evaluate reactions to individual data bases. As a result, it has not been possible to extract any findings that are directly relevant to CAIN.

The Georgia Centre did not commit themselves to an opinion as to whether CAIN users were satisfied or not since there is virtually no quantitative evidence to support any firm views on this matter. Their general impression is that CAIN does give satisfactory results for agricultural scientists working in the applied fields or in cases where profiles do not require extensive coverage of peripheral subject areas. For many profiles, CAIN is combined with other tapes such as Biological Abstracts to provide full coverage of the user's subject field, but there are some cases when profiles are searched only against CAIN.

B.5 USER: AGRICULTURAL RESEARCH SERVICE, Unites States Department of Agriculture (USDA), Beltsville, Maryland 20705

B.5.1 SOURCE OF DATA: (i) Visit with H.D. Burton at the Agricultural Research Service, September 1973. (ii) Report entitled "A user-dependent SDI system - They said it couldn't be done" by H.D. Burton, paper presented to the Special Libraries Association, Annual Meeting, Pittsburgh, Pennsylvania, June, 1973.

B.5.2 CAIN SEARCH EXPERIENCE: The Data Systems Application Division of the Agricultural Research Service (ARS) have been operating an SDI service free of charge to U.S. Department of Agriculture (USDA) users since May, 1972. The SDI service provides searches of the following five bibliographic data bases: Biological Abstracts (BA), BioResearch Index (BRI), Chemical Abstracts Condensates (CAC), CAIN, and Engineering Index (Compendex) for a user community of about 1000 scientists and engineers. These users are located either in the immediate Beltsville area or in different regions of the United States where they are employed at one of five major regional laboratories or at small experimental stations. Several advisory or extension services also make use of the ARS SDI System but they represent a very small percentage of the total user group.

Although the National Agricultural Library (NAL) and the ARS are both a part of the U.S. Department of Agriculture, there is no link between the NAL system for producing CAIN and the ARS SDI system which incorporates CAIN as one of its data bases. The ARS subscribe to the NAL tape CAIN service on comparable terms as those offered to other organisations and their application of CAIN for computer-based searches is independent of any of the search facilities provided within the NAL.

The ARS are currently running about 6000 SDI profiles on behalf of just over 1000 users but the average number of profiles per user is closer to three than to six. The distortion results from several users having 20 to 30 profiles where the user might be a librarian attempting to monitor the major areas of research for an entire laboratory. The number of profiles (and corresponding users) searched against each of the five data bases is summarised below:

Data Base	No. of profiles	No. of users
CAIN	1041	570
BA	1519	809
BRI	1482	796
CAC	1825	1016
Compendex	289	199
	6156	

As can be seen from the above table, CAIN does not emerge as the dominant data base as far as usage is concerned with there being only 600 CAIN users compared with about 1000 CAC users and 800 BA users. One can conclude from this that in general, USDA scientists find their profiles to be better suited to BA and CAC than to CAIN, although there has been no formal monitoring of user preferences.

As a general policy, CAIN is always searched in combination with other tape services such as BA and CAC since it is felt that CAIN on its own, would not give comprehensive coverage of the user's field of interest. However, it is also generally accepted that these other tape services are likely to be lacking references that CAIN will have and therefore they too need to be supplemented for complete coverage. Just recently, a retrospective search facility has been introduced and 100 searches of the full CAIN file of around half a million references have been carried out to date. It is considered that to be ecomonically practical, the batch size for a retrospective search run must be at least 10 queries and based on this figure, ARS expect to be able to run retrospective

searches at two-weekly intervals. From search experience gained thus far with CAIN, it appears that retrospective searches are likely to average at about 25 per month, indicating substantial user interest in this search application.

B.5.3 PROFILE CONSTRUCTION: With the ARS system, the user is expected to develop his own profile and initiate all revisions. He is provided with a "User's Guide" outlining procedures for formulating profiles and suggesting sources for identifying useful terms such as thesauri, dictionaries and classification codes. Although some advantages are claimed for this "user self-reliance" such as lower running costs and greater responsiveness between the user and the system, it is generally accepted that some user constructed profiles are bound to be weak, and without the presence of an interface between the user and the system it will not always be possible to identify the changes that are needed. The current practise of relying on the user to develop his own profile has been justified mainly for reasons of cost and lack of trained staff within ARS to handle this function.

In constructing a profile to be searched against CAIN, users normally rely almost entirely on their own knowledge of the subject area and are not as a rule encouraged to consult any thesauri or printed services such as Bibliography of Agriculture to aid in the selection of suitable search terms. No comparative searches are attempted with other services as a means of identifying any weaknesses in the CAIN profile. As a consequence there is little opportunity to assess the extent to which CAIN search results are being adversely affected by omissions in a profile rather than by weaknesses inherent in the data base.

Compared with tape services such as BA and CAC which provide structured term vocabularies, CAIN is considered to require greater effort from the user in identifying sets of word variants and synonyms to define his subject area and is therefore felt to be less convenient to use. A consequence of this, is that in cases where the task of constructing a CAIN profile involves a series of tedious and complicated look-up procedures, the user may not be able or willing to devote sufficient time to this activity with the result that the profile will be an inadequate expression of his field of interest.

B.5.4 USER SEARCH SYSTEM: The ARS computer search system is a modified version of software developed by the University of Georgia, known as the UGA Text Search System. Details of this search system are given in Section B.4.4.

B.5.5 USER ASSESSMENT OF CAIN: User reaction to CAIN is mixed. Biological Abstracts is generally preferred over CAIN but users consider that CAIN is still needed for its broad coverage of the agricultural literature. The coverage requirements of USDA users were thought to be more heavily slanted towards the pure research and agriculturally related subjects than is normal at most organisations and this may partially explain why users find their profiles better suited to BA and CAC than to CAIN.

No formal monitoring of the performance of any of the tape services has been attempted by ARS. In an effort to assess user benefits being derived from the SDI service, a questionnaire survey was carried out in 1973. Unfortunately, the information collected referred to the overall service and was not specifically related to any particular data base. As a result, it has

not been possible to deduce any user reactions that are directly applicable to CAIN from this study. The overall findings of the survey indicate that 75% of users are satisfied with the SDI service and that time savings and better coverage of their subject fields were the principal benefits claimed.

No data has been collected on CAIN's search performance or user reactions to it, but the general impression is that 75% of the total number of CAIN users find the service helpful and of this group, 25% are very satisfied with their results. In general, users react more favourably to the CAIN retrospective search facility than to the CAIN SDI service. CAIN is considered to perform particularly well in retrospective searches, providing a relatively high percentage of relevant hits for many of the questions posed. It is thought that for both SDI and retrospective searches, CAIN gives a higher percentage of relevant hits than does BA but that BA outperforms CAIN in providing better recall.

The Agricultural Research Service are not entirely satisfied with CAIN but still continue to run more CAIN profiles than any other organisation covered in this study. Although CAIN is considered to be weak in several areas, they do recognise CAIN to be a necessary component of their information service becauseof its broad coverage of the agricultural literature. It is claimed that the extensive use of CAIN by USDA scientists is due not to strong user satisfaction with CAIN but to the realisation that preferred services such as BA and CAC need to be supplemented by CAIN if complete coverage of the user's subject area is to be achieved. However, it must be noted that this view is based on a subjective judgment of the entire user group and is not supported by any statistical evidence of individual user reactions to the different tape services.

B.6 USER: SENTRAL for FORSOKSMETODIKK og DATABEHANDLING, (FDB-sentralen), Vollebekk, Norway

B.6.1 SOURCE OF DATA: (i) Visit with E. Eik at FDB-Sentralen, August 1973. (ii) Letter from E. Eik, May 2nd, 1973. (iii) Report entitled "Appreciation of a bibliography on magnetic tape" by Egil Eik, FDB-sentralen, May, 1973.

B.6.2 CAIN SEARCH EXPERIENCE: FDB-sentralen (Central Office for Search Methods and Data Processing), a computer centre, have been operating a CAIN SDI service since October, 1971 in co-operation with the Library of the Agricultural College in Vollebekk for research scientists based at the College. They also run SDI services using Food Science and Technology Abstracts (FSTA) and World Textile Abstracts (WTA) for other government Research Institutes but neither of these tapes is searched in combination with CAIN. Largely satisfactory use of the Bibliography of Agriculture printed service led the College to its initial choice of CAIN for SDI purposes.

For the first 16 months, the CAIN SDI service was offered at no cost to users, but since January 1972, users have been charged at the rate of about £14 per profile per year. When the service became operational in October 1971, there were 120 SDI profiles (corresponding to about 100 users) and by 1972, when the charging policy was introduced, the number of profiles

dropped to around 70. During 1973, the number of profiles has diminished even further and at present only 39 profiles are currently being processed by the centre.

FDB-sentralen have, at no time, attempted to use CAIN for retrospective searches although the requirement for such searches may well exist. They consider that CAIN would be unsuitable for this purpose because of the lack of controlled index terms for items in the file. In the case of the FSTA and WTA tapes which do provide controlled subject terms with each item, they are maintaining cumulative files stored on disc for retrospective searching.

Throughout 1971, users were requested to keep a formal record of the relevance of hits obtained to allow a subsequent analysis of CAIN's performance during this period. The evaluation covered a total of 69 profiles which during 1971 were being searched against the 1970 and 1971 CAIN tapes. The findings are tabulated below.

Profile No.	No. of hits	No. of relevant hits	
		Major interest	Secondary interest
1	12	4	3
2	896	81	72
3	175	10	28
4	180	16	22
5	475	47	40
6	4	2	2
7	18	4	1
8	840	223	197
9	154	26	36
10	402	16	77
11	386	14	51
12	147	7	12
13	27	3	4
14	30	8	10
15	15	5	4
16	14	11	1
17	129	5	10
18	48	14	8
19	337	24	34
20	268	213	38
21	1	1	0
22	4	0	1
23	198	17	73
24	37	12	9
25	35	3	6
26	248	21	13
27	60	22	10
28	16	10	1
29	54	20	11
30	146	39	60
31	114	28	54
32	27	3	1
33	265	86	104
34	95	44	12
35	202	33	93
36	184	76	86
37	258	22	79
38	516	12	75

(Cont.) Profile No.	No. of hits	No. of relevant hits	
		Major interest	Secondary interest
39	3	3	0
40	64	22	9
41	1275	101	321
42	852	152	81
43	735	79	60
44	91	5	1
45	628	69	37
46	497	63	80
47	13	1	11
48	227	12	6
49	93	7	7
50	84	41	23
51	15	4	6
52	12	10	1
53	146	14	29
54	155	49	15
55	44	0	1
56	76	8	11
57	1546	42	102
58	30	17	5
59	37	2	4
60	48	11	25
61	427	28	41
62	0	0	0
63	162	25	90
64	81	5	16
65	264	30	48
66	34	7	0
67	163	31	0
68	29	8	0
69	7	0	1

A total of 14,845 hits were retrieved during this period of which 2028 (14%) were of prime importance and 2369 (16%) were found to be of secondary interest. These results indicate that even when items classified as being of secondary interest are taken into account, a fairly low percentage (30%) of the hits received by users were relevant to the questions posed. Only about one third of the profiles being searched produced search results of which over 50% were relevant to the particular user concerned, while for a similar portion of the profiles (just under one third) a relevance of less than 20% was observed for the answers received.

B.6.3 PROFILE CONSTRUCTION: Users are required to construct their own profiles with only limited guidance from the staff at the Computer Centre. FDB-sentralen recognise the advantage in having an information specialist trained in the vocabulary conventions of a data base to interact with users during the formulation of profiles, but have not as yet allocated staff for this purpose. As a result, the user is more or less left on his own to select appropriate search terms relying almost entirely on his general knowledge of a particular subject. No crosschecking of alternative printed services is attempted during the initial search runs, and hence there is very little opportunity to assess whether poor search performance is indicative of a weak profile or the failure of CAIN to provide

adequate coverage of the users' field of interest. On average, profiles contain about 10 search terms where a term can be entered in full or as a word root.

B.6.4 USER SEARCH SYSTEM: FDB-sentralen have written their own software for processing user profiles against the serial CAIN file. The following fields in the CAIN records are searched: subject category codes, language and translation codes, personal and corporate authors, subject terms, journal titles and document titles. The search logic is based on the standard AND, OR, BUT NOT logical operators. The search system runs on an IBM 360/40 with 64K and programs are written in Assembler.

The software as it is designed at present, limits the user to a maximum of 25 search terms and does not permit the use of parenthesis (i.e. nested logic) in the Boolean search term statement. For example, the system could not handle the question: A OR (B AND C) OR D. Several optimisation features are incorporated in the search program to allow rapid scanning of the textual fields in the CAIN data base. A typical search is run in 30 mins. lapse time (charged at £1.30/min.) and the average batch comprises 40 profiles resulting in a running cost of roughly £1 per profile per run.

B.6.5 ASSESSMENT OF CAIN: The general impression is that CAIN provides inadequate coverage of the agricultural literature and its performance as measured by the relevance of the search results is poor. The detailed and extensive evaluation of the percentage of relevant hits obtained during 1971 certainly demonstrates a low standard of performance for the service but the extent to which other factors such as badly constructed profiles or software limitations may have contributed to these poor results has not been ascertained.

The poor performance of CAIN throughout this period was one of the factors which led FDB-sentralen to undertake a questionnaire survey of user reactions to CAIN. Three categories of user were distinguished: (A) CAIN users, (B) scientists who had completely stopped using the CAIN service and (C) scientists who had never submitted a CAIN search profile. The results of the survey carried out in March 1973 are summarised below under each user group studied:

Group A (questionnaire sent to 20 users and all responded)

Questions	Answers	
(a) Why do you use CAIN?		
(i) to follow your field of interest	19	(95%)
(ii) to alert yourself to current information on a new subject	4	(20%)
(b) Are you satisfied with CAIN?		
(i) Yes	16	(80%)
(ii) No	4	(20%)
(c) How useful is CAIN to you in your subject field?		
(i) Very	7	(35%)
(ii) Moderately	12	(60%)
(iii) Slightly	1	(5%)

Group B (questionnaires sent to 14 scientists with a 79% response)

All asked the question: "Why did you stop using CAIN?"

A selection of the replies received appears below.

(i) It doesn't cover major scientific journals very well and it includes too many popular, non-scientific articles.

(ii) It did not give very satisfactory coverage of my field of interest.

(iii) Very few of my relevant hits were of interest since I had already obtained these references from other sources.

(iv) Very few of my hits were relevant.

(v) I could not afford the service.

(vi) Too many of my references were inaccessible because of foreign language problems.

(vii) I found it difficult to assess relevance from titles alone.

(viii) I did not have any confidence in a service which relies mainly on titles for subject searching.

Group C (questionnaires sent to 23 scientists with an 82% response)

Questions	Answers
Why have you never used CAIN?	
(i) It does not cover my subject field very well.	7 (30%)
(ii) I cannot afford the search charges	5 (22%)
(iii) I am sufficiently satisfied with my present sources of current information	9 (40%)
(iv) I was unaware that such a service was available	12 (52%)

The finding that most CAIN users were satisfied with their results is in keeping with the realisation that users will only continue to pay for a service if it is of measurable benefit to them. It is noteworthy that 52% of the scientists contacted who had never tried CAIN gave as their reason that they had not been aware that the service was available.

In an attempt to evaluate the type of material being covered by CAIN, FDB-sentralen selected articles from all Norwegian and Northern community (Scandinavian and Finnish) journals in a 12 month portion of the CAIN data base (August 1971 to July 1972). A total of 1472 articles were located and subsequently analysed. In general, the findings indicate that some scientific journals fail to get complete article coverage, implying that items of major importance may have been overlooked and that a substantial portion of the total articles covered, originate from non-scientific, advisory journals which are not likely to be of interest to their users.

B.7 USER: ALBERTA INFORMATION RETRIEVAL ASSOCIATION, Alberta, Edmonton, Canada

B.7.1 SOURCE OF DATA: Letter from A. FitzPatrick, May, 1973.

B.7.2 CAIN SEARCH EXPERIENCE: The Alberta Information Retrieval Association have had two years experience in running retrospective searches of CAIN and presently handle between 100 and 150 searches per year. Users are required to pay for searches but these charges have only discouraged users with a casual or border-line interest in the service,

B.7.3 USER SEARCH SYSTEM: No information.

B.7.4 USER ASSESSMENT OF CAIN: User reaction to CAIN is generally good.

B.8 USER: UNIVERSITY OF CALIFORNIA, Campus Computing Network, Los Angeles, California 90024

B.8.1 SOURCE OF DATA: (i) Report entitled "CAIN PROJECT - Phase II Final report" by E.C. Jestes, November, 1972. (ii) Letter from B. Briggs, to Dr. U. Schutzsack (IFIS), Dec.21st, 1972. (iii) Letter from B. Briggs, August, 1973.

B.8.2 CAIN SEARCH EXPERIENCE: The University of California have been running SDI searches using CAIN since January, 1972, and as of December, 1972 were searching 126 profiles representing 117 users. Unfortunately, it has not been possible to locate details of the current usage of CAIN at the University of California in the time available for preparing this report.

The increase in the number of profiles run against CAIN since the service was first introduced can be observed in the followin table:

Month	No. of CAIN profiles	No. of users
Jan. 1972	56	56
February	77	70
March	88	81
April	97	90
May	105	94
June	108	102
⋮	⋮	⋮
Dec. 1972	126	117

Quantitative studies of the relevance of the CAIN search results have been carried out throughout 1972. Their findings indicated that on average, 70% of the hits obtained from CAIN were relevant to the particular questions posed.

B.8.3. USER SEARCH SYSTEM: CAIN searches are currently carried out using the IBM TEXT-PAC system but this software is to be replaced shortly by a new in-house search system.

B.8.4 USER ASSESSMENT OF CAIN: User reaction to CAIN is generally favourable. A questionnaire survey was carried out in 1972 to assess user benefits in terms of time saved as a result of using CAIN, compared with the costs of operating CAIN. The two main findings to emerge from this cost benefit study were:

(i) the majority of their CAIN users considered the service to be of material value (i.e. time saved, duplication avoided), and
(ii) the reported benefits in terms of time saved by researchers far outweighed the overall costs of providing the service.

B.9 USER: NATIONAL RESEARCH COUNCIL OF CANADA, National Science Library, Ottawa, Canada

B.9.1 SOURCE OF DATA: Letter from G. Mauerhoff, September, 1973.

B.9.2 CAIN SEARCH EXPERIENCE: CAIN has been run on an experimental basis only thus far, but the National Research Council (NRC) is now preparing to make CAIN available across Canada through their CAN/SDI service. The CAN/SDI (Canadian Selective Dissemination of Information) system is in its fourth year of operation and it presently incorporates around 14 commercial tape services, providing SDI searches for a user community of about 4000. The CAIN SDI service will be made available through the Library of Agriculture, Canada, which has been designated as the CAN/SDI centre responsible for user contacts in the agricultural sciences.

B.9.3 USER SEARCH SYSTEM: The National Research Council have developed a general purpose IR package for searching the multiple data bases included in their service.

B.9.4 USER ASSESSMENT OF CAIN: Since CAIN is not yet operational, it has not been possible to report on user reactions. Experience to date has been confined to a number of experimental SDI searches with the objective of familiarising the NRC staff with the structure and content of the CAIN file. The NRC are convinced of the need to provide access to CAIN in order to obtain the necessary coverage of the agricultural literature.

B.10 USER: CAIN/CAB PROJECT GROUP, UKCIS, The University, Nottingham.

B.10.1 SOURCE OF DATA: (i) CAIN Progress Reports Nos.1, 2, 3A, 4 (Aug.1971 - June 1973), CAIN Project Advisory Committee Papers (March 1972 and Sept. 1972) and Annual Report for 1972. (ii) Visit to UKCIS, April 1973. (iii) Letter from J. White, October 1973.

B.10.2 CAIN SEARCH EXPERIENCE: The CAIN/CAB Project Group are planning a study of the use of the CAIN tapes for running SDI searches but have not as yet acquired any search experience with CAIN. Their main interest in CAIN to date has been in studying and comparing the content relative to agricultural and agriculturally-related data bases.

A comparative study of the completeness and speed of coverage achieved by the CAB printed journals and the CAIN tapes was initiated in August 1971. A computer-based system was developed for matching serials and non-serials in both data bases with the former compared by journal or condensed title and citation and the latter by authors and citation. Standardisation routines have been devised to allow a consistent representation of citation, title and author fields in both data bases to reduce errors during the matching process. Partial matches will be displayed in order to permit visual checking in the event that identical entries had been inconsistently or incorrectly recorded.

The results of a small scale study indicate that the subject overlap between CAB and CAIN is roughly 40%. The difference in timeliness (as expressed by the difference in nominal publication dates) of articles common to both was found to be fairly small. The general trend seems to suggest that items normally appear in CAIN one to two months in advance of their publication in CAB.

The study has been extended to include the BIOSIS data base (selections from Biological Abstracts and BioResearch Index) with samples from CAIN, CAB and BIOSIS being matched. This work has not yet been completed, but early results indicate that only a small percentage (10 to 15%) of the BIOSIS items are to be found in the CAB and CAIN sample files. The low number of matches was expected since the BIOSIS data base covers agriculture within the broad subject field of the life sciences rather than as its primary subject. From the results obtained thus far in the sample matching of CAIN and CAB against BIOSIS, the Project Group estimate that about 22% of items in CAIN are present in BIOSIS and that around 32% of articles in CAB are to be found in BIOSIS. Current studies of the relative timeliness of the data bases seem to indicate that articles common to BIOSIS and CAIN generally appear in BIOSIS two to three months prior to their notification in CAIN.

B.10.3 USER SEARCH SYSTEM: **The feasibility of the UKCIS INFIRS system for running CAIN SDI searches has been demonstrated.**

B.10.4 USER ASSESSMENT OF CAIN: The points raised by the CAIN Project Group and summarised below, relate to their detailed study of the various fields in the tape for matching purposes (and not to search experience):

(i) The document type code does not clearly designate document type in all cases since the identical code can be used for both journal articles as well as government reports, conferences and annual reports.

(ii) In several thousand CAIN records processed to date, it is observed that the secondary category code field is effectively not being used (i.e. blank or identical with the primary code).

(iii) It has been noted that in a substantial number of cases, an inconsistent format has been adopted for recording citation data (e.g. p.1-22 vs 22p.). Few inconsistencies have been observed in studies of the author fields.

B.11 USER: CHEMIE INFORMATION UND DOKUMENTATION BERLIN, 1000 Berlin 12, Steinplatz 2.

B.11.1 SOURCE OF DATA: Report entitled "NAL-CAIN-Tape", April 14th, 1972 (8p.).

B.11.2 CAIN SEARCH EXPERIENCE: No details of any search experience are recorded. CID Berlin are not regular subscribers to the CAIN service and it appears that the tape was studied on a "one time" only basis.

B.11.3 USER SEARCH SYSTEM: No information.

B.11.4 USER ASSESSMENT OF CAIN: CID Berlin have studied the frequency of occurrence of the various "directory" data elements in the CAIN tape records. Their study was based on a sample of about 5000 records extracted from the January 1972 tape and the results are summarised below.

Data elements (variable no. of fixed length segments	Percentage occurrence of the data elements in each document
	%
New Book Shelf descriptive information	0
Document Titles	100
Personal Author names	95
Corporate authors	5,3
Author Biographic data	0
Journal title	100
Pagination	99
Document Date	100
NAL Call No.	100
Subject Terms	6,8
Notes	34
Information to patents ...	0
Series Statement	41
Abstract/Extract	0
Tracings	2
(for later use)	0
Non-vocabulary cross references	0,8

CID Berlin are critical of the number of data element types which are either not being used at all (e.g. abstract) or are only being used rarely (e.g. subject terms). The use of fixed length segments (65 bytes) for the different data elements is a further criticism since it results in tape storage space being wasted for small data elements (e.g. documents date and NAL Call No.). It is further noted that for large data elements such as titles it is necessary to reconstruct the original statement by combining the different segments, adding another processing step each time the item is read.

Finally, they consider that two subject codes are insufficient and suggest that further subject classifications should be included. It is noted that in the tape sample studied, only one subject category code had been assigned to each item.

B.12 OTHER CAIN SUBSCRIBERS: The name of the remaining subscriber for whom no information has been located is:

Dr. Carlos Cuadra,
System Development Corporation,
2500 Colorado Avenue,
Santa Monica, California 90406.

APPENDIX C - OTHER CAIN USER STUDIES

C.1 SWEDISH STUDY

An analysis showing for each of the CAIN subject category codes, the number of documents that have been assigned that code has been completed by the Swedish Agricultural College. The results listed below refer to the 1972 CAIN tapes:

	CAIN subject categories	No.of documents/ category
0505	General agriculture & rural sociology	2563
1005	General agricultural economics & land economics	1266
1010	Agricultural administration & management	2021
1015	Agricultural production costs & returns	1430
1020	Agricultural production distribution	1982
1025	Statistical data & methodology	523
1030	Outlook, policies, programs & legislation	1895
1505	Consumer protection	2105
1510	Human nutrition	1410
1515	Home economics	172
2005	Agricultural products - general	2173
2010	Dairy products	1448
2015	Livestock products	1070
2020	Poultry products	261
2025	Field crop products	2094
2030	Horticultural products	1831
2035	Feed products	671
2505	Animal husbandry	1644
2510	Livestock biology	4526
2515	Livestock feeding	4533
2520	Livestock breeding	3519
3005	Veterinary medicine	1948
3010	Infectious & parasitic diseases	6404
3015	Non-infectious diseases	1463
3020	Miscellaneous diseases	720
3505	Forestry - general	1173
3510	Forest management	1378
3515	Silviculture	1234
3520	Forest industries	2746
4005	General plant science	873
4010	Plant taxonomy & geography	4219
4015	Plant ecology	1185
4020	Plant morphology, anatomy & cytology	2532
4025	Plant genetics & breeding	3258
4030	Plant physiology & biochemistry - general	7498
4035	Physiology & biochemistry of field crops	4427
4040	Physiology & biochemistry of horticultural crops	3584
4045	Physiology & biochemistry of forest trees	1028
4050	Field crops - culture	3872
4055	Horticultural crops - culture	3690
4060	Miscellaneous economic plants - culture	388
4505	Plant fungus diseases & control	3479
4510	Plant bacterial diseases & control	399
4515	Plant virus diseases & control	1278
4520	Miscellaneous plant diseases, injuries & control	2233
4525	Weeds & herbicides	2068
4530	Insect pests & controls - general & miscellaneous plants	1227

4535	Insect pests & controls - field crops	1317
4540	Insect pests & controls - horticultural crops	1309
4545	Insect pests & controls - forest trees, products	738
4550	Insect pests & control - products	316
4555	Insect pests & control - animal and man	1045
4560	Pesticides - general	1589
5005	General entomology	6149
5010	Taxonomic entomology	3470
5015	Apiculture & sericulture	824
5505	Agricultural engineering	1152
5510	Farm equipment	2278
6005	Soil science	4421
6010	Soil improvement materials	3069
6015	Soil resources & management	959
6020	Water resources & management	2128
6505	General natural resources & environmental pollution	2085
7005	Life sciences	1947
7505	Physical sciences & mathematics	779
8005	Chemistry	744
8505	Technology	328
9005	Economics & administration	1899
9505	Social sciences & humanities	1006
9705	Information science	371
		143364

C.2 NORWEGIAN STUDY:

The Norwegian Agricultural College report the following percentage distribution of CAIN references under the 20 main subject groups for the Jan.72, Jan.73 and Feb.73 CAIN tapes:

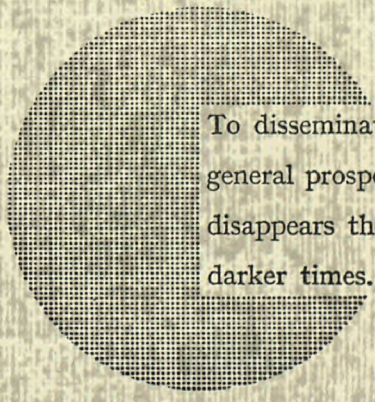
	Main subject classifications	%age distribution
05	General agricultural and rural sociology	1.5
10	Agricultural economics	7.0
15	Consumer products	2.6
20	Agricultural products	6.8
25	Animal Science	9.8
30	Veterinary medicine	7.0
35	Forestry	4.7
40	Plant Science	24.3
45	Plant diseases insect pests and control	12.3
50	Entomology	7.8
55	Agricultural engineering	2.4
60	Natural resources management	7.3
65	Environmental pollution	1.6
70	Life Science	1.6
75	Physical science & mathematics	0.4
80	Chemistry	0.5
85	Technology	0.2
90	Economics and administration	1.3
95	Social science and humanities	0.6
97	Information Science	0.3
		100.0%

****

NOTICE TO THE READER

All scientific and technical reports published by the Commission of the European Communities are announced in the monthly periodical **"euro-abstracts".** For subscription (1 year: B.Fr. 1 025,—) or free specimen copies please write to:

Office for Official Publications
of the European Communities
Boîte postale 1003
Luxembourg
(Grand-Duchy of Luxembourg)



To disseminate knowledge is to disseminate prosperity — I mean general prosperity and not individual riches — and with prosperity disappears the greater part of the evil which is our heritage from darker times.

Alfred Nobel

SALES OFFICES

The Office for Official Publications sells all documents published by the Commission of the European Communities at the addresses listed below, at the price given on cover. When ordering, specify clearly the exact reference and the title of the document.

UNITED KINGDOM

H.M. Stationery Office
P.O. Box 569
London S.E. 1 — Tel. 01-928 69 77, ext. 365

BELGIUM

Moniteur belge — Belgisch Staatsblad
Rue de Louvain 40-42 — Leuvenseweg 40-42
1000 Bruxelles — 1000 Brussel — Tel. 12 00 26
CCP 50-80 — Postgiro 50-80

Agency :
Librairie européenne — Europese Boekhandel
Rue de la Loi 244 — Wetstraat 244
1040 Bruxelles — 1040 Brussel

DENMARK

J.H. Schultz — Boghandel
Møntergade 19
DK 1116 København K — Tel. 14 11 95

FRANCE

Service de vente en France des publications des Communautés européennes — Journal officiel
26, rue Desaix — 75 732 Paris - Cédex 15e
Tel. (1) 306 51 00 — CCP Paris 23-96

GERMANY (FR)

Verlag Bundesanzeiger
5 Köln 1 — Postfach 108 006
Tel. (0221) 21 03 48
Telex: Anzeiger Bonn 08 882 595
Postscheckkonto 834 00 Köln

GRAND DUCHY OF LUXEMBOURG

Office for Official Publications of the European Communities
Boîte postale 1003 — Luxembourg
Tel. 4 79 41 — CCP 191-90
Compte courant bancaire: BIL 8-109/6003/200

IRELAND

Stationery Office — The Controller
Beggar's Bush
Dublin 4 — Tel. 6 54 01

ITALY

Libreria dello Stato
Piazza G. Verdi 10
00198 Roma — Tel. (6) 85 08
CCP 1/2640

NETHERLANDS

Staatsdrukkerij- en uitgeverijbedrijf
Christoffel Plantijnstraat
's-Gravenhage — Tel. (070) 81 45 11
Postgiro 42 53 00

UNITED STATES OF AMERICA

European Community Information Service
2100 M Street, N.W.
Suite 707
Washington, D.C. 20 037 — Tel. 296 51 31

SWITZERLAND

Librairie Payot
6, rue Grenus
1211 Genève — Tel. 31 89 50
CCP 12-236 Genève

SWEDEN

Librairie C.E. Fritze
2 Fredsgatan
Stockholm 16
Post Giro 193, Bank Giro 73/4015

SPAIN

Libreria Mundi-Prensa
Castello 37
Madrid 1 — Tel. 275 51 31

OTHER COUNTRIES

Office for Official Publications of the European Communities
Boîte postale 1003 — Luxembourg
Tel. 4 79 41 — CCP 191-90
Compte courant bancaire: BIL 8-109/6003/300

OFFICE FOR OFFICIAL PUBLICATIONS OF THE EUROPEAN COMMUNITIES
Boîte postale 1003 — Luxembourg 6564

CDNA05169ENC