

Text Retrieval at the UK National Audit Office



Auditors require quick and easy access to an increasingly complex range of information. Mandy Dolphin highlights the text retrieval initiatives in the UK Audit Office.

Mandy Dolphin



Mandy Dolphin joined the NAO in 1988. She now manages a busy Library providing a wide range of information-related IT systems such as the Central Reference Database, Bulletin Board and Library Automation.

Background

The UK National Audit Office (NAO) is responsible for providing independent assurance, information and advice to Parliament. This covers the proper accounting for Central Government expenditure, revenue and assets and the economy, efficiency and effectiveness of the use of resources by Government departments.

The NAO employs 780 staff. The majority are based in the London headquarters in Victoria. The remainder are based in a number of locations around the United Kingdom and in Rome and Geneva.

Three quarters of the NAO staff are professionally qualified accountants or under training. The remainder provide administrative and specialised support such as information and computer services. The NAO also engages consultants and other staff on short term contracts where specialist knowledge and skills are required.

Why the NAO considered Text Retrieval

As with other audit and accountancy institutions the NAO makes use of a growing body of professional guidance and precedent. Searching and identifying items can be time consuming if carried out manually and expensive in terms of staff time. More often than not the information is needed immediately for it to be of any benefit. The NAO also wanted to make savings on the number of publications issued to the office. With the development of a Text Retrieval system it was envisaged that fewer publications would need to be

issued on a one per section basis saving costs.

Text Retrieval was one of a number of strategic systems identified in an Information Technology (IT) strategy carried out in 1987. Prior to this the NAO had purchased a variety of computer equipment and developed a number of stand-alone systems. These ranged from PCs used by secretaries for wordprocessing to specialist time recording and personnel systems.

The Text Retrieval package selected was BRS/Search which offers information management at the personal, department and corporate level. The key reasons for selecting BRS/Search were that it has a low storage overhead; is easy to tailor; is feature rich; could easily be integrated with word processing software; and it has mini and micro computer versions.

The IT Strategy identified two main types of internal reference databases;

- Common or Central databases which were relevant to all staff; and
- Unit or specialist databases, relevant to small groups of staff.

We developed the central databases first and introduced our Central Reference Database (CRD) in November 1989.

We currently have twenty central databases operational. Some of these are security restricted databases, eg for Directors and Management Board. The rest contain reference material such as copies of our reports to Parliament, our audit manual and guidance and internal management circulars. These are available to the 721 staff who have access to the NAO's Office Automation computer system.

Central Reference Database: Available Databases

CN00 PAC Reports 1969 - 1983
CN01 NAO Reports, PAC Reports and Treasury Minutes
CA01 CN00 and CN01 combined
CN02 Finance Manual
CN03 Audit (including Audit Manual and Auditing Standards)
CN04 Index to Select Committee Reports
CN05 Accounting (including Accounting Standards)
CN06 Strategic Plans: Study Proposals
CN07 Personnel
CN08 Management Circulars
CN09 Travel
CN10 Office Services Guide
CN11 Audit Precedent
DC01 Directors Circulars
MB01 Management Board Minutes
MB02 Management Board Papers
MB03 MB01 and MB02 combined
UA01 Audit Advice
UA02 CN11 and UA01 combined
UA03 National Audit Office Act
UA04 Policy Unit Indexes and Abstracts
UA05 UA03 and UA04 combined
UA06 Policy Unit Enquiry Forms
UE01 Defence Working Papers

In 1991 we began developing Unit databases and we now have a total of eight. One contains the full text of enquiry forms used to record audit advice given by our technical section; it can be searched each time a new enquiry occurs to ensure consistent advice. A similar database is now in operation for Policy Unit enquiries.

One advantage of the BRS/Search retrieval software has been the ability to concatenate databases, that is to allow searching of more than one database at a time. We currently provide this facility in three areas and plan to expand its use over the next year.

Making Searching Easier

In 1989 when we first introduced CRD, we used a BRS interface known as SearchMate. This interface was customised to integrate with Uniplex, our Office Automation software, in three ways; to save a retrieved document into a Uniplex Word Processing file; to Print using Uniplex utilities; and to access other Uniplex functions such as electronic mail while retaining the search session in BRS.

Whilst the SearchMate interface was well used by staff accessing the databases frequently, the occasional user found it

frustrating and difficult to remember the necessary steps needed to retrieve relevant information.

In 1993 we therefore introduced an alternative interface, developed by Simdell Ltd. This new front-end is based on Uniplex. Searching is carried out by filling in a Uniplex form and all commands are executed by using function keys. Where possible these function keys replicate the function keys of the Uniplex package to provide consistency.

The introduction of the new interface has proved successful. Occasional users feel more comfortable with an interface which has the "look and feel" of the Word Processing package with which they are familiar. Access statistics indicate that the new interface has increased usage of CRD by some 30%. The more "expert" users do however continue to use the original SearchMate interface.

Taking Text Retrieval further

The extent to which data should be computerised depends upon the retrieval requirements and the "importance" of the information. We began with full text retrieval; because the main central

reference documents are of a permanent nature, requiring limited amendment and provide information which needs to be available to a large number of people.

The second phase, which has now started, tackles the problem of providing effective, yet efficient retrieval of a wide range of important information which is constantly being added to and is of a semi-permanent nature.

Without some structured method of retrieving references to files or the information contained within them it is likely that either the information will be lost or it will be retrieved too late.

We identified four ways of handling this material, and we have developed a cascade approach to its retrieval:

- **File level** - the database contains a list of file titles and reference numbers only
- **Index level** - the database contains the file titles and an index to all the papers held within each file
- **Abstracts and Keywords** - the database contains an abstract and assigned keywords to indicate the subject content of files
- **The Full Text** - the database contains the full text of all files.

The Pilot Study

In order to test this approach and give a clear understanding of the issues, problems and costs involved we decided to conduct a pilot project in our Policy Unit.

From the outset we recognised that significant resources would be required to assign abstracts and keywords and to reproduce full text of files; this made it likely that when the facility was offered more widely users would chose to start with file level and index level retrieval. In the longer term however, upgrading to abstract, keyword and full text was probable for files of corporate interest.

Computerising a mess can only result in a worse mess. We therefore undertook a careful review of our hard copy files before we computerised them. This involved checking for duplication, providing the files with meaningful titles, and weeding old material.

With full text retrieval an index is produced by the retrieval software. However as we were dealing with files and index sheets rather than full text, the first step was to produce an index for the Policy Unit area.

Constructing a good index is the key to effective information retrieval although it frequently gets little attention as it is laborious and time consuming. Once the index was produced it was used to construct a classification scheme. The scheme is based upon the Dewey Decimal Classification scheme and is used to assign file numbers for location purposes; the numbers also reflect the subject content of the files.

Pilot Databases

For the pilot study we developed a file tracking system and three databases, at index, abstract and keyword and full text levels, we also offered a concatenated database covering all three.

For retrieval at **file level** we have developed a file tracking system using the Uniplex Office Automation Package. This enables retrieval by title or classification number and also provides a file history to establish who currently holds a file and who has seen it previously. Using this system it is also possible to produce an up to date file register in hard copy.

The **index level database** provides references to the NAO's audit precedent files. Each file is a separate document with a title and reference number together with an index to all the enclosures held in that particular file.

The **abstract level database** covers central policy and is an extension to the index database. For each enclosure on a file a mini abstract is provided. Each file and thus each document within the database also has a subject classification number which can be searched, and which also acts as the file location point when retrieving the hard copy.

The **full text database** provides access to material related to the NAO Act, such as the first draft of the Bill, Adjournment debates and relevant Select Committee Reports.

Within each document we have embedded cross references which link documents within the database. This approach is a fore-runner to the development of hypertext searching which we hope to introduce during 1994. Hypertext enables the user to highlight an embedded reference and view the relevant document immediately rather than having to search for it.

Use of External Text Retrieval

Text Retrieval is more than just an in-house tool. There are several thousand

commercially produced reference databases which can be accessed for a fee. Our auditors need access to wide ranging and vast amounts of information relating mainly to value for money investigations. These can range from developments in the financial control of the National Health Service; to texts on coastal protection; to material on Urban Development Corporations; to the Management of military clothing.

Accessing this information manually, using indexes or abstracts is slow and time consuming, if you can get access to the material. But using external databases is a highly effective way of providing a wide range of relevant material at relatively low cost.

In the NAO we make use of eight major hosts - the largest, Dialog, offers over 500 databases. There is an overlap between many of these hosts, so searching can be complicated.

We also make use of databases available on CD-ROM. This can save costs because there is no on-line connect charge. However updating will often be done only monthly or quarterly compared to daily or weekly with on-line systems.

Each of these databases has different searching characteristics, and searching can be quite sophisticated. The complexities and access costs have led us to regard use of external databases as a job for information professionals.

We undertake many small searches on an ad hoc basis. In 1990 we launched a Research Service embracing both internal and external sources of information to provide a comprehensive pack of information to VFM Research teams embarking on new studies. The Research Service has proved extremely valuable particularly in saving audit teams time and in providing the most up to date information available.

A New Medium

Over the past three years there has been a growing interest by other UK Government departments and SAI's in gaining electronic access to the NAO's reports. In 1993 we therefore began to examine the possibilities of providing our reports in CD-ROM format.

We approached BRS Software Products (now BRS Dataware) to produce the CD-ROM for us. Because we were already using BRS Software on our central Unix system there were minimal problems in

providing our data for inclusion in a CD-ROM.

BRS designed an interface for us based on their Epi-Find software which is used in 95% of BRS's CD-ROM production. The interface was tailored to incorporate the search fields we required for retrieving relevant reports.

The CD-ROM contains a suite of databases providing access to NAO Reports, PAC, PAC Reports and Treasury Minutes from 1969 onwards. The CD-ROM is self-contained with full loading instructions in an accompanying booklet, and full on-line help once the CD-ROM is installed.

A copy of the CD-ROM was recently distributed to each English-speaking SAI. The CD-ROM will be updated on a six monthly basis. If your SAI has not received a copy and would like one, please contact the UK NAO Information Centre Help Desk on + 44- (0)171-798 7264.

Now that we have embarked on producing our own data in CD-ROM format we plan to investigate other types of information suitable for this medium. We could, for example, provide our audit teams on local audit with CD-ROMs containing our Audit Manual. Auditors would then be able to access a large volume of data when they are with clients, without having the security problems associated with dialling into a central system.

Moving Towards Document Management

The NAO generates a vast amount of information in paper and magnetic form and we also have a large archive. But to be honest our archive management is poor, and reviewing/weeding old files is the last job on anyone's list.

But our IT strategy, particularly the text production and filing parts of UNIPLEX, and BRS Search offer the start of a total document management system. And the introduction of document image processing (DIP) systems - although not yet fully cost-effective - are likely in future to enhance our records management.

So we are investigating options for scanning text, image and graphics and making them accessible to end users with database software and a high resolution monitor. Information can be stored on optical discs held in a "jukebox" and terminals can be networked. Links could also be set up with information held on CD-ROM and microfiche to provide a total document management system.