

# 1 Background, history and aims

## 1.1 The project proposal

The project “Microprocessor Networking in Libraries” was conceived in 1980. The proposers, Neil McLean and Mel Collier, had long been involved in library automation, and were currently engaged in implementing a large multi-site telecommunication network at the Polytechnic of Central London, linked to the SWALCAP library data processing cooperative [1]. They had also worked on early developments in microcomputer-based applications, and were increasingly interested in the gulf between received opinion about the capability of microcomputers for substantial library applications and the apparent, but largely untested, potential of the newer technology.

The catalyst for this thinking was the appearance on the British market in 1980-81 of microcomputer networks. Microcomputer networks are the low cost end of local area network (LAN) technology, and were the first off-the-shelf products to result from LAN research stretching back into the early 1970s. The appearance of microcomputer networks was attended by unprecedented publicity (even for the computer press) focusing on the predicted importance of this technology for the much heralded “office automation revolution”.

It seemed to the proposers that a revolution in office organisation and associated information management was of itself of prime importance to the library world: many information handling operations in libraries are no different in principle from general office tasks. It would be particularly useful to show whether micro-networks are appropriate not only for general office tasks but also for specialised library data management and information retrieval. The key question for this smaller scale technology was how it would perform in relation to very large textual files and numerous users.

A proposal was therefore made to the British Library Research and Development Department to investigate microcomputer networks for

library applications [1]. While the researchers would report generally on this topic, it was decided to focus on the online public access catalogue (OPAC) as a trial application. This choice was initially made on the grounds that, although there were already many computerised cataloguing systems, the catalogue itself “remained a hard copy [product] with its attendant problems of non-currency, inflexibility and recurrent cost”; it was felt that the flexibility of network configurations might help to reduce the cost of implementation of OPACs, particularly for smaller libraries, and that the project might also attempt to tackle some of the problems of the limitations of MARC records for users’ access to bibliographic information [2, Section 4.4].

The proposal gained British Library Research and Development Department support, and also attracted funds from the Department of Industry (now Trade and Industry) as an associated project with the Information Technology Centre (now Library Technology Centre), which was also based at the Polytechnic in Regent Street. The two projects officially began in June 1982, with the first project staff being appointed later in the year.

## **1.2 Staffing**

Stephen Walker was appointed as research fellow in November 1982 and Gillian Venner, also as research fellow, in January 1983. Nathalie Mitev was appointed as research officer in July 1983.

A small amount of additional programming assistance was available for a few months in 1984.

Because the course of the research was influenced strongly by the interests and aptitudes of individual team members, it is appropriate to say something about the background of each of the workers.

Nathalie Mitev is an information scientist, trained in France although she has worked a good deal in the UK, including working on a previous British Library Research and Development project. She has been involved in setting up information systems, as a search intermediary, and has some programming experience. Her interests lie mainly in the area of the study of users and the use of information systems, in which she is currently researching for a PhD.

Gillian Venner completed a physics and maths degree before becoming a computer programmer. She gradually came to specialise in library applications, first at Bath University, where she developed the Minicatalogue with Philip Bryant, and then at Dundee University, where she was responsible for the library automation.

Stephen Walker has been a mathematician and statistician and spent many years in industry. In recent years he has moved into information science, and his main interest is in information systems for non-specialist users. He has designed and written several information retrieval (IR) systems.

Throughout this report the word “team” usually refers to one or more of these three people.

### **1.3 History of the project**

By the start of the project (November 1982) the proposers had already acquired a LAN — a “Cluster/One” system from Zynar (later Nestar) — for demonstration purposes in the Polytechnic’s Information Technology Centre. The researchers satisfied themselves that it would be possible to implement a (possibly rather limited) online catalogue system on this network, given suitable software.

It had been hoped by the proposers that an existing text retrieval or database management package could be bought or leased, and substantially modified, if necessary, so that it would operate within the LAN environment, and allow adequate access to a catalogue file.

There were at that time few IR systems for microcomputers capable of supporting reasonably large files and providing suitable access facilities for an online catalogue, and this is probably still the case. The team investigated two IR packages, both micro versions of established text-retrieval systems, and attended demonstrations of two database management systems (DBMS). It seemed unlikely that the DBMS would be suitable on the grounds that, whatever the suppliers may say, they are not very suitable for storage of, and retrieval from, variable length records of largely textual content. As for the IR systems, only one of them might have been operable on the Apple II computers which had been obtained for the Cluster/One network as user stations. This system appeared to have a very poor response time, and the suppliers were not very encouraging when asked whether it would be possible to make the source

code and documentation available so that it could be made suitable for catalogue access. There were at that time no LANs available which could handle the emerging 16-bit IR packages such as the micro versions of CAIRS or BRS.

Thus the only way to proceed was for the team to set about developing its own access software, as well as software to produce a suitable source file from MARC records on an exchange tape and to transfer it onto the network's disc store.

In April 1983 an additional grant from the British Library Research and Development Department made possible the purchase of a more advanced LAN system, a "PLAN 4000" from the same supplier. It was installed, with five terminals (or stations), in the Regent Street building of the Polytechnic. Such software as had been written was transferred to the new network and all further development was carried out on it. The new LAN gave somewhat faster disc access and, more importantly, the possibility of mounting a catalogue file of quite substantial size. This meant that it became possible to test the software and hardware under realistic conditions: with at least the complete holdings of a site library of the Polytechnic. It should be noted that the higher cost of the PLAN network (Section 3.3.1) moved the balance of the project towards systems aimed at somewhat larger libraries than had been intended.

LOCAS supplied a MARC exchange tape of the 30,000 records for items held in the largest site library, and the team wrote programs to enable the file to be read and partially stripped on a mainframe and transferred to the network ready for further processing and inversion. By July 1983 the file had been transferred to the PLAN 4000 network.

To get some idea of the likely performance of the system some 6000 records were mounted under an existing IR system (*FIRS II*, a later version of the system described in [3]); this demonstrated that the hardware would probably behave adequately under a realistic load, but this system was not tested on "real" users as the user interface is not suitable for untrained people. A few hundred records were also loaded in a format suitable for the software package *Cardbox* (Caxton Software Ltd).

In April 1984 the Polytechnic bought a second PLAN 4000 network with ten stations. This was installed in the Riding House Street site library and was to be used for library purposes other than the present project as well as for the live evaluation of the OPAC.

By the spring of 1984 the team had written an IR system, without any more in the way of a user interface than was needed for debugging and testing, of a fairly conventional inverted index type but designed so as to minimise the amount of disc access. Preliminary work had been done on record and index display layouts and modes of user input.

Tests showed that the system would probably perform adequately with a larger file than had been envisaged, so the team carried out a second transfer onto the network of the entire Polytechnic union catalogue of some 90,000 records, and also took the opportunity to include more of the MARC fields than had originally been intended, and to redesign the OPAC's bibliographic record.

A stand had been booked at a conference and exhibition "Public access to library files" to be held by the Centre for Catalogue Research in September 1984. Spurred by this goal, the team produced a working OPAC system with about one day to spare, and took it with a 33,000 record subfile and four user stations to Bath. The system was named *Okapi* (if this has to be an acronym it stands for "Online keyword access to public information"). *Okapi* proved far more successful than had been hoped. At that time there was no automatic logging of the system, but the team obtained, by observation of, and chatting to, hundreds of users (librarians, students, people in the library automation business) at the exhibition, many useful comments and suggestions.

The final part of the project was some evaluation of the system as an OPAC. It had always been intended that it should be set up in one of the Polytechnic's site libraries as an alternative to, or replacement for, the microform catalogue. It had also been intended to do some testing under controlled conditions. Through the Centre for Catalogue Research, Linda Reynolds was retained as consultant to help with the design of evaluation methods.

During the autumn of 1984 evaluation procedures were drawn up. These consisted of (1) fairly comprehensive self-logging within the search programs, (2) user observation sheets and (3) a questionnaire to be administered by one of the team. It became apparent that there would not be time to carry out the controlled experiments.

In November 1984 a terminal was installed in the Riding House Street site for library staff use, and at the end of November one public access terminal was set up and data collection was started. As expected, the

OPAC proved rather popular with the users (mainly undergraduate students of social sciences and humanities). Detailed protocols of some seventy user sessions were collected, as well as extensive machine logging of several thousand searches.

#### 1.4 Aims of the project

It can be seen that what started as work primarily concerned with the investigation of the suitability of a particular type of computer configuration for library automation soon became a study of the public access catalogue. As work progressed, it became apparent to the team that learning how to make good OPACs is probably the most important aspect in the application of computers to libraries: the “data processing” side of library automation is hardly an area for further research, and is not of much direct benefit to users, but the OPAC potentially offers greatly improved access to library materials. The OPAC can also be seen as a prototype of interactive retrieval systems for general users — something which viewdata systems might be if television transmission standards were such as to allow good response times and adequate display of information. (That the study of catalogue use was important for the design of “universal” IR systems for general users was realised at least fifteen years ago: the much-cited study by Tagliacozzo and Kochen, *Information-seeking behavior of catalog users* [4], arose out of a project for “learning to design a ‘growing encyclopedia system’ ”.)

Evaluation of the hardware took second place, but some features of the prototype OPAC, which was one of the results of this research, do take advantage of facilities which are more readily available with networked hardware than with multi-user single-processor configurations.

#### References

- 1 **Collier M W** and **Piper D**. Multi-site networking: experience of the Polytechnic of Central London. *Program* 18 (2), 1984, p147-156.
- 2 **Collier M W** and **McLean N F**. *Microprocessor networking in libraries*. Research proposal to the British Library Research and Development Department. April 1982.
- 3 **Walker S**. An information retrieval package for microcomputers. *Program* 16 (3), 1982, p171-179.
- 4 **Tagliacozzo R** and **Kochen M**. Information-seeking behavior of catalog users. *Information Storage and Retrieval* 6 (5), 1970, p363-381.