The Archival Electronic Record in the Nineties : Recent Initiatives at the National Archives of Canada

by **MAGGIE MACLELLAND**

Introduction

The Government Archives Division (GAD) at the National Archives of Canada has extensive experience with archival records in electronic form, beginning in the early 1970s. Over the past five years the division has refined its approach to appraisal, emphasising the functions revealed in the records regardless of their medium. The approach has resulted in a steady increase in the acquisition of archival records in electronic form. This article describes how the acquisition function has changed to meet the challenge presented by these records.

Until 1990 the function of acquiring archival electronic records remained basically unchanged in the division, and concentrated on survey data from agency or departmental main-frame computers. The inexorable spread of automation in government departments and agencies, however, means that a growing percentage of the documents they create are in electronic form. Divisional management realized the need for a new approach that would cope with the increasing quantity and diversity of these documents. In late 1989 a proposal was made for an in-house facility that would satisfy the need.

The Setting

The Government Archives Division was formed in 1986 by the merger of the Federal Archives Division and the Machine Readable Archives Division. It is responsible for the appraisal of all unpublished textual, micrographic, and machine-readable data created by the departments and agencies of the Government of Canada, and for acquiring and describing that portion of the records having enduring archival value.

At the time of the proposal, while the Archives had plans to install a department-wide office support system, GAD was one of the few divisions in the department with a fully functioning Local Area Network (LAN). Divisional access to the MINISIS databases of finding aids and item-level descriptions was by way of either low-speed modem from a few microcomputers on the LAN, or printout volumes shelved in the reference room. For the archival electronic records themselves, a service bureau provided main-frame facilities for verification, processing, copying, access, and distribution, at an annual cost of almost $400,000. A small group of specialists in the division prepared the data and dealt with the more esoteric aspects of formats, structures, character sets, etc.

The Proposal

The goal was to establish an in-house facility for assessing, acquiring, processing, and distributing archival electronic records, with the following provisions:

I. Must be fun to use for archivists and researchers

Using any computer (with the possible exception of the Macintosh) takes a significant amount of training and experience. The Government Archives Division needs to deal with archival electronic records from main-frames, minis, and networked and stand-alone personal computers. It is virtually a full-time job for anyone with a knack for dealing with computers to keep up to date technically in any one of these areas. It is unlikely, therefore, that more than a few archivists would willingly deal with naked data as it comes from the donor agency. Moreover, the investment in time makes this impractical as it detracts from the archivists' other functions of appraisal, description, and public service. Researchers should be able to do their work with electronic records on a "self-serve" basis, much as they do now with paper documents. This was certainly not a possibility with the service bureau facilities.

The proposed in-house facility, therefore, had to provide a way for archivists and researchers to get at electronic documents in a way that minimized the need for training. It had to provide an environment that enabled people to work with a minimum of effort. This means an interface that supports several software tools simultaneously, while minimizing the learning required to use them. The simplest solution would be a graphical user interface (e.g. the Macintosh computer "desktop"), with icons to point at and click, drag, and drop. Actual data access requires a suite of software tools to handle traditional databases and large collections of text.

II. Must not exceed the cost of existing facilities and services, including service bureau use

Using the existing budget as an upper limit minimized the administrative complications. The strategy was to reduce charges at the service bureau and direct the resulting funds to the purchase of hardware and software.

III. Should take advantage of existing and emerging government standards and National Archives facilities

Meeting this criterion should reduce the amount of work required to acquire, preserve, and distribute archival electronic records, especially considering the long periods of time over which the National Archives will have to maintain their accessibility. It also supports the federal government's standards initiatives, which in turn encourages federal agencies to adopt the standards, thus making acquisition a simpler task.

IV. Must be vendor independent

This is a key factor in the long-term maintenance of archival electronic records. Vendors come and go, change formats, and drop applications. Such activities would jeopardize the value of National Archives holdings unless those holdings are independent of any particular vendor.

V. Must have virtually unlimited, incremental growth potential

Both the quantity of data acquired and the demands for access will grow continually. It would therefore be unwise to develop a system with an inherent limitation on its size. The traditional approach of completely replacing an existing system with one of a larger capacity is too expensive, both in hardware and data conversion costs. A system that can accommodate the addition of storage and processors on an incremental basis is far more economical and flexible, and can adapt to new uses or a growth in demand for services.

The Project

Where possible, we planned to use the division's existing hardware, such as the LAN workstations, for the in-house facility, and to borrow whatever we could to assemble a prototype. The idea was to "try before you buy"; to be satisfied that the concepts were sound. It would require time and effort, but little if any capital investment.

Based on the decisions about the prototype, the next step was to purchase the necessary hardware and software. Pending their delivery, the prototype continued in use as a learning tool. Implementing the system depended to a large degree on this acquisition phase. A post-implementation review measured the success of the project against the goal criteria, and set the guidelines for future development.

The proposed in-house facility would be smaller than the division's existing LAN and cheaper than the contract with the service bureau. Its impact on the way in which the National Archives handles archival electronic records is another matter. For it to be successful, the in-house facility had to integrate with other departmental projects: the Archives' new office support system, the archival holdings database, EDP systems in other divisions, and facilities in other branches of the Archives. To accomplish such integration, the first step was to ensure that system design and implementation of the facility did not exclude connection to and integration with the other systems. This meant that any design decision would have to be made with knowledge of the other systems' designs. It did not mean that all of these systems were to become parts of one mega-project.

Four people in the division were performing archival electronic records-related duties using the main-frame facility of the service bureau. These people formed the nucleus of a small technical group responsible for implementing and operating the in-house facility. By January 1991, GAD had successfully migrated its data from the service bureau facility and its main-frame environment, and had demonstrated that archivist access and use of the system is practical, with comparatively minimal training needed.

The in-house facility uses computers from three vendors, operating systems from two vendors, and peripherals and software from suppliers too numerous to list. All of these products are commercially available and off-the-shelf, with one exception: some utilities written in-house for the nine-track tape operations. Access from divisional desktop workstations to both the in-house facility and the MINISIS databases is provided over the departmental office support system (Banyan VINES).

Current work is concentrated on improving divisional access to reference materials—such as the Record Group inventory descriptions and existing MINISIS databases—using text-search and database access tools, and to providing working tools for text preparation. Work began in May 1991 and is in the final, evaluation stage. A volunteer group of ten people from the division each had up to a half-day of familiarization and training for On-line Reference; in most cases this was sufficient. The User Group is evaluating the On-line Reference application. Experience with it to date has ranged from its demonstration by archivists for the International Congress of Archives (Montreal, September 1992) to its use in response to telephone enquiries.

Future direction could extend the On-line Reference application across the division, and to other divisions; and could make it locally and remotely available to researchers. The division is currently experimenting with accessing selected archival electronic records through On-line Reference internally, and has recently joined the Ottawa FreeNet as an information provider.

The in-house facility, with its On-line Reference application, won the 1992-93 Award of Excellence in innovation from SCOAP, the Ottawa-based Society of Canadian Office Automation Professionals. SCOAP cited in particular the application's potential for significant organization-wide impact on the provision of information for National Archives staff and clients.

The On-line Reference Application

Selected Record Group inventories, their associated MINISIS finding aid databases, and the electronic version of the General Guide series for the Government Archives Division (GAD Guide)—comprising almost 400 megabytes (the equivalent of between 200,000 and 400,000 pages) of information—are available on the in-house facility. On-line reference enquiry is done from some workstations in the division.

A user need only select an area of interest and ask a question. At the workstation, an archivist opens a "reference" window, which connects through the office support system (the Banyan VINES network) to the UNIX network of the facility. The archivist then logs in to the UNIX host machine and brings up the On-line Reference application in a windowing environment. Clicking the mouse on an icon representing a file invokes the search tool (Research Assistant) with the file loaded and ready to process enquiries.

The graphical user interface was designed to resemble the way archivists currently organize their information. A simple three-level directory structure was made to lead from the top (most general) to the bottom (most specific) level of information. Icons were created that roughly approximate these levels (e.g., three-ring binders, filing cabinets, magnetic tapes). The actual information, which exists in word processing, desktop publishing, and HP3000 format, was converted to ASCII and indexed using Research Assistant.

A Few Words about Costs

Since the project had to be accomplished within the existing budget, expenditures for main-frame processing had to be reduced. By mid-1990, it had become evident that the administrative overhead of the Service Bureau operations was excessive. It was not possible to eliminate this altogether immediately, because other National Archives users were in the midst of projects that could not be taken off the main-frame economically. GAD therefore negotiated agreements with three other users providing for elimination of the need for the service bureau by the next year (1991-92). The other divisions now use a DOS-based system for their nine-track tape operations, and the Service Bureau contract was not renewed.

Some typical costs, given here as an indication of the order of magnitude of expenditure encountered in setting up the facility, were:

$62,000 - UNIX network
$11,000 - Research Assistant (5 DOS licenses, 5 UNIX licenses)
$10,000 - Overland 9-track tape facility

The cost for the On-line Reference application's hardware and software is $21,680, broken down as follows:

Host:

$7,000 80486 PC with 1 GByte HDD
$1,700 UNIX with Looking Glass
$1,200 additional Looking Glass licenses
$5,000 Research Assistant licenses

$14,900

Per workstation:

$200 RAM
$210 DESQview/X
$155 TCP/IP extensions

$565 x 12 = $6,780

In summary, from a budget of over $400,000 in 1989-90, the facility now operates on a budget of less than $100,000 per annum.

Contribution of the In-house Facility

The migration from the service bureau to the in-house facility has created a departmentally-controlled computing environment for archival electronic records that is practical for archivists to work with, at a substantially reduced operating cost. The division has benefitted from the savings through the acquisition of upgrades to desktop workstations and selected software packages.

A separate organizational unit, the Electronic Systems Projects Division (which since the spring of 1992 has been part of the Informatics and Records Services Branch) looks after system planning, administration, operation, training, processing of records, and evaluation of hardware and software for GAD. This means that archivists can now concentrate on appraisal, description, and public service. The in-house facility is more responsive to their needs, and it provides them with a suite of applications that they can use to access the records with minimal effort. Its flexibility enables it to handle traditional databases as well as enormous quantities of text. The division now has a wealth of reference material in electronic text form, in a variety of formats, accessible by various means. On-line Reference gathers the information into a single format and makes it accessible through a single interface at the desktop workstation level.

GAD archivists and ESPD project officers now regularly appraise electronic records in office systems, in document imaging systems, and in integrated main-frame/mini/pc networks, on media ranging from punched paper tape to magnetic-optical disc.

ESPD, through the in-house facility, equips the archivist with the appropriate tools, technical advice, support, and training needed to cope with the electronic record as an archival item. ESPD verifies the data and documentation for accuracy and completeness, provides estimates of the costs of acquisition, processing, and distribution, prepares technical specifications that form part of the transfer agreement with the donor agency, and handles the physical transfer of the records, their conversion to standard forms and formats, and their distribution.

The archivist is able to access, view, and manipulate the records for purposes of appraisal and description and for responding to researcher enquiries. In the course of appraisal an archivist can use the in-house facility to "pre-view" a sampling of records obtained from the donor agency on almost any of the media mentioned above. Once the records have been acquired, archivists produce working copies, which may then be loaded as databases, run through data analysis, or indexed for text searching—whatever technique and software are most appropriate. More than half the archivists in the division are now using the MINISIS databases on-line, rather than the printed versions, to answer enquiries—quickly locating information additional to that found in the hardcopy material. Access requests that could have taken days of manual searching can now be done in hours.

Although much has been done to enable the division to deal with archival electronic records, access by archivists and researchers is a continuing challenge. With the technical support provided by the in-house facility, archivists in the division are now in a better position to develop archival practices that take into account the requirements of the electronic record in the 1990s.

Technical addendum

The Platform

The POSIX-compliant UNIX System V, Release 3, is the chosen operating system for the in-house facility. Inter-machine communications are handled by TCP/IP over the existing Ethernet cabling in the division. The Banyan-to-UNIX connection uses TCP/IP. Network management of the facility uses Network File System (NFS) and Remote Processing Control (RPC), and the user interface resides on the X Windows graphical environment. The facility has a relational database management system incorporating SQL.

The significant characteristic of this platform is its independence from specific vendors, hardware, and software. It can use a range of relatively cheap microcomputers (e.g. 80386 EISA), accessible from the divisional LAN workstations, which can be configured to meet a variety of needs. Off-the-shelf hardware and software minimize costs. Distributed processing enables the system to cope with tasks that would exceed the capacity of any single machine, while permitting incremental changes, additions, and upgrades to parts of the system without necessitating complete system replacement.

Acceptable X Windows/Motif products are now available; the in-house facility uses one of them (Looking Glass). The division acquired Research Assistant for text-searching. Its user interface is not entirely satisfactory, primarily because of some rather bizarre data display. Database access can be done through Ingres (included with the UNIX product) for those who prefer SQL, or FoxBASE+ for those who are more comfortable with dBase.

Configuration

A client/server configuration was set up so that all the facility's files and software are resident on a UNIX platform in the Electronic Systems Projects Division (ESPD), which functions as the X Clients—while each member of the user group in GAD has a software package on the local hard drive that allows the DOS workstation to function as an X Server.

The configuration for GAD's user group workstations is:

Intel 386 with 4 MB RAM (Departmental standard workstation)
Additional 4 MB RAM
Minimum 50 MB HDD
Mouse
MS-DOS 5.0
Connection to Banyan VINES Office Support System
PC/TCP Ver. 2.05
TCP/IP network option
Either DESQview/X Ver. 1.02 multitasking, windowing, graphical DOS operating environment (X Server), or
HCL-eXceed/W 3.2.0 (X Server) and MS Windows 3.1

On the UNIX side in ESPD, the in-house facility is configured as follows:

8 UNIX Hosts - Intel 386 & 486 with up to 64 MB RAM
7.63 GB storage
Overland 9-track tape processing facility with Tapeview
TCP/IP
UNIX network gateway to Banyan VINES LAN
Interactive UNIX System V/386 Release 3.2
Motif X Window Manager
Looking Glass Ver.2 graphical user environment
Research Assistant (Beta version for UNIX) natural language text search and extraction tool

Outstanding technical issues

1. The interaction between DESQview/X Version 1.02 and Banyan is not stable; as far as can be seen, this is not the case when DESQview/X is communicating only with the UNIX networking software. Another X Server, HCL-eXceed/W, is being evaluated in the division. Early indications are that it may be more stable than DESQview/X in its interaction with Banyan. On a per workstation basis it is more expensive and requires the use of MS Windows, which DESQview/X does not. However, unlike DESQview/X, it could be installed as a network application on Banyan.
2. The Beta for UNIX version of the search engine, Research Assistant, does not do a lot of the functions which its DOS version performs. The software vendor (Readware) is re-writing the UNIX version in C, but it is difficult to predict if the work will ever be completed. ESPD is therefore looking into possible alternatives to this search engine.
3. Neither Ingres nor FoxBASE+ meets the criterion of ease of use for accessing conventional databases. The search, therefore, continues.
4. The user group members continue to evaluate the On-line Reference application. They are optimistic that, once the workstation end has been stabilized, more use will be made of the application. Some would like to have more data to search on. Other archivists are interested in the application; we are not adding more copies of the X Servers at this stage, but user accounts can be made available for interested archivists. One impediment to extending its use to more people is the limited number of Looking Glass licenses: only seven concurrent users are supported by the floating license server.