Does Practice make Perfect?

The Electronic Records Program
at the National Archives of Canada
1973-2003

Yvette Hackett
E-Archiving for Posterity
Antwerp
June 26, 2003

Introduction

I've been asked today to offer the perspective of an archives with an active program to acquire and preserve records in electronic form. This request comes at an interesting time, since the National Archives of Canada will shortly cease to exist. There is currently legislation before the Canadian Parliament designed to merge the National Archives and the National Library of Canada into a new and "innovative knowledge institution" called the Library and Archives of Canada. It is too early yet to understand in any detail how this new institution will be structured or will function, but it does mean that today, in discussing the electronic records program at the National Archives, I speak only about the past.

The electronic records program has formally existed since 1973 and during those 30 years, it has known great changes and more than a few ups and downs.

Types of electronic documents

In any discussion of electronic records, it is first important to categorize them as they are not a single, homogenous type of document which can be approached in a single, homogenous way. What is appropriate for data generated in a database environment is not necessarily directly applicable to textual records or digital audio files.

When the Machine Readable Archives Division (which I will refer to as the MRA) was created at the National Archives, automation had not yet come to "traditional" records. There were no desktop computers, off-the-shelf word processing softwares or e-mail programs. What the Machine Readable Archives set out to preserve was "data" - surveys, opinion polls, scientific data and social science research data sets from both government and private sources. Historically, the National Archives would have acquired the reports written about the data; analysis and recommendations based on the data; policy documents developed as a result of the data; but not the data itself. But in 1973, the potential archival value of these data sets was acknowledged and became part of the MRA's acquisition mandate.

Today, we are more focused on a second type of record - textual records created in electronic form, which include all types of more traditional records such as correspondence, reports, minutes, etc.

Appraisal

When appraising all types of electronic records, I think the first question we have learned to ask at the National Archives is "What is the paper equivalent?" Most computers are used to create documents which, in the past, were created on paper. When determining the long-term value of documents, this question can often help crystallize the archival perspective.

The second, general question which must be asked relates to life-cycle. We are finding that the life-cycle of electronic records is somewhat different than for paper records, though each of the traditional stages is still there. In the mainframe era, concerns about technological obsolescence had pushed acquisition into the active stage of a record's life. Today, we have returned to a more traditional approach - the acquisition of inactive records. Though it seems that every year, somebody will argue that, given the easy replicability of digital records, there is no reason why everybody can't have a copy. This suggestion happily ignores all cost issues relating to the maintenance of significant numbers of duplicates, and the far more serious problem of whose copy would be the authoritative one.

Where appraisal is concerned, we have found that the same criteria applied to paper records can be successfully applied to textual records in electronic form.

Data

Where data is concerned, more specific issues must also be addressed. For example, many of the early records management, correspondence tracking, and file retrieval systems, are not archival records, but finding aids. They contain only a small portion of the relevant information recorded elsewhere in paper files and are used primarily to track workflow, measure performance and provide quick responses to client inquiries about the status of their file. Some of the content of these databases can be loaded into an Archives' descriptive system, but should not be mistakenly acquired as electronic records simply because they are digital.

Some large "data" applications can best be compared to case files, a long-standing problem for archives in their paper form, due to their large extent. The fact that they have been automated (or far more frequently even today, partially automated) has not solved the problem. It is only in the last 2 to 3 years that we have started to see fully electronic case file systems. In each case, these required a significant “imaging” component to convert the many paper records still being generated by the process, whether it be military personnel files or patent applications. When the creator wants to destroy all paper records immediately after scanning, and when their lawyers have agreed with them that this is an appropriate action, then the archivist can seriously examine the electronic record as the authoritative record of that institution. But even at this point in the automation process, you will likely find a system where quality control is inadequate and too little attention has been paid to the stability of the records beyond the institution’s “active” time frame.

Some operational databases might be of interest as research data but in this scenario, one must understand first that there is a tremendous amount of duplication of information among the databases of major federal programs. Some form of overall selection is still required. Though we no longer acquire data primarily for its statistical research value, our appraisal questions today still reflect, to some extent, this earlier mandate. Some databases are better than others because:

- they contain a great deal more information about the individuals or "things" than others
- they contain a larger sample of Canadians than others
- they were used to deliver programs that were more significant than others

In recent years, this has led us to concentrate on large, national systems such as the census records collected by Statistics Canada, and the tax records collected by Revenue Canada. Databases in other departments are evaluated not only against record-keeping requirements internal to the department, but also against the "key" federal databases which we have already identified for transfer.

Also, portions of many federal databases are passed from department to department, with various additions and deletions occurring with each transfer. In these cases, appraisal extends to measure the various "value-added" processes in each department and taking the record only at the point where it contains a maximum of valuable, non-duplicated information.

One pitfall to avoid when appraising large data applications is to lend too much archival weight to the fact the data can be easily manipulated, one attribute absent from paper case files. Manipulability does not equal archival value. Some data will simply not have real long-term value. House-keeping records remain house-keeping records, even in digital form.

Another set of important questions that need to be asked when database applications are being appraised relate to documentation, that information which sets out how data in an automated application is organized, and how it is to be interpreted. This can range from basic record layouts to sampling methodologies to algorithms used to correct data variances. A database with little or no documentation may be technically readable, but it is also undecipherable.

Available documentation must also be comprehensive. Information technology personnel are notorious for maintaining only what they currently need to run a system. But a description of a database as it exists in 1997 may be useless if you are trying to decode a 1974 history file, especially if it is primarily numerical.

Textual records in electronic form

For more than 20 years, textual records in electronic form, eventually including e-mail, have been created on mainframes, micro-computers, local and wide-area networks. Despite the numerous configurations, they all support the creation of what we, as archivists, generally recognize as our traditional records. The National Archives was a little slow in issuing a statement that e-mail can constitute a record. Quite frankly, it had seemed rather self-evident to us, given the type of information which was being conveyed via e-mail, even within our own institution.

When appraising this type of record, it is important to understand two things:
1) the records management culture of the institution, as expressed in its policies and procedures and in the extent to

which they are observed by employees; and
2) the “automation period” during which the electronic records were created. For a significant portion of time after its installation on my desktop, my computer was no more than a big typewriter. While I typed on it and saved documents that I was working on - first on diskettes, then on my hard drive and finally into a personal space on the network - anything important, anything official, anything that I wanted to sign even as a courtesy, and anything that I wanted to circulate, was dutifully printed to paper.

Unstructured textual records in electronic form, even more than unstructured paper records, require significant resources if they are to be brought under any kind of reasonable intellectual and physical control. As with the "manipulability" issue discussed earlier, the first attribute of electronic office systems to be extremely wary of is "free text retrieval". It is a useful tool if you are really stuck, but it in no way provides an adequate replacement for the context which is provided by proper arrangement and a standardized hierarchical description with standardized access points.

And we are, in my opinion, just coming out of the unstructured era. Individual employees' caches of diskettes, the content of their hard drives or of their individual work spaces on network servers are analogous to those personal filing cabinets found in every public servant's office - a random selection of documents, ranging from the significant to the silly, organized according to personal and often eccentric classification systems which are only irregularly applied. If your archives does not search out this type of paper record, do not suddenly start to accept them in electronic form.

Our experience has consistently shown that the key to textual records in electronic form is, as with paper, proper classification. During appraisal, the archivist must first understand how the creating institution controls its records. Does it still maintain a "print to paper" policy? This is not as rare as you might think, even today and it is most likely to be true for the 1 to 2% of government records currently identified for long-term preservation at the Archives.

Is there still a central records office? Has the organization introduced an electronic records management system which allows the application of a coherent file classification system to its electronic records? Has it actually been fully deployed in the institution or are there only small areas involved in pilot projects? And the most recent question that I’ve learned to ask - if the institution does have the software, is it actually being used to control electronic records or are they still limiting its application to the control of paper records? From an operational perspective, what does the organization depend on as the "record" of its activities? Ask them what they would take to court.

The answers to these questions will begin to define the potential value of any office system and the degree of duplication which may exist between the paper and electronic components of their records. As I pointed out earlier about case files, very few organizations are truly operating in a paperless environment. Far more common is a hybrid situation, with most activities generating part paper, part electronic files, with potentially neither system offering a comprehensive record. Again, it is only in the last few years that we have seen the ratio shift from a preponderance of paper to significant amounts of electronic records among the current records being created in departments. It will be years before most of this material is ready for transfer to the Archives. And again, as with case files, except in rare circumstances, the only way to get a 100% electronic file is to introduce scanning in a fairly significant way.

Our most famous foray into textual electronic records was our earliest - the records of the Trade Negotiations Office, which were acquired in 1989. The TNO negotiated Canada’s first free trade agreement with the United States, an extremely important event in our history, the benefits of which are still being debated today. That experience taught us a great deal about the pitfalls of unstructured office systems, as well as the problems associated with acquiring from a backup, rather than from a live system. It is not an accident that we have not attempted this type of acquisition again.

Our study of the Trade Negotiations Office records suggested that almost 30% of the electronic records were not duplicated in paper files - a significant proportion. It's important to remember, however, the corporate culture of the TNO - a short-lived organization with a high staff turnover and too much money to devote to quick technological solutions in support of a high-visibility, high-pressure mandate. Would a similar study of later records produce the

same results?

Interestingly enough, that opportunity presented itself in the mid-1990's with the records received from the Royal Commission on Aboriginal Peoples. Some time after the Royal Commission had issued its final report and the paper records had been transferred to the Archives, the archivist responsible for the fonds received a phone call. Approximately 60 diskettes had been found in the offices of the Privy Council Office, the central agency of the federal government which administers the creation of Royal Commissions in Canada. Nobody there knew what was on them, nobody there had worked on the Royal Commission - could they not just send them to us since we already had the rest of the records. First, we asked for a list, hoping to prevent any actual transfer to the Archives and all the subsequent paperwork that that entails. Unfortunately, the list of diskette labels was not conclusive and we requested transfer of the diskettes themselves. Almost immediately, 1,000 diskettes arrived - I guess they "found" more after receiving permission to ship. Except for the 60 diskettes which contained the text and graphics from the final published report of the Commission, the balance of the diskettes were labeled with file classification numbers. These matched the paper files already in our holdings. We felt it offered an interesting opportunity to compare the state of record-keeping practices with those of the Trade Negotiations Office. While with TNO we had found that almost 30% of the electronic records were not duplicated in paper files, this time we found less than a 2% difference. Clearly, in the 10 years between the two special agencies, some lessons about records management in a hybrid environment had been learned.

Within the National Archives, I frequently argue against the acquisition of textual records in electronic form if they are also available on paper. I should specify that this is properly classified and filed paper. My opinion has evolved slowly over years of attempting to process too many poorly arranged, poorly documented and poorly maintained transfers of electronic records from the "big typewriter" period - since these are the records Archives are still most likely to be receiving.

I suspect one of the great fallacies in the electronic records debate is that everything somehow gets easier, cheaper and faster if you are presented with records in electronic form. While costs relating to physical storage are generally lower than for paper, the crucial calculations actually relate to the time it takes archivists to appraise, technically evaluate, monitor, process and describe electronic records when compared to paper. The underlying technological dependence of these records means that, for electronic records, these costs will be even higher than for those traditionally associated with art, photography and especially audio-visual records.

Currently available Information, Document and/or Records' management applications are just now developing into good, standardized tools which should help automated office environments stabilize and begin to resemble their traditional paper counterpart - the central records' office. (But employees are largely resisting their use.) Content Management and Digital Asset Management systems seem to have decided to start from scratch, emphasizing their support for tracking sales or royalty payments rather than records management requirements. Last year, we began to closely examine RDIMS - the Record, Document and Information Management System which federal government departments are strongly encouraged to implement in Canada. This solution involves the integration of a number of software packages, including a document management one and a number of records management applications to choose from. Buried in one of the software manuals, we found a configuration switch that would automatically destroy the earliest draft of a document when it's 5th version was filed. Another example was a switch which, if improperly set, deleted the document history when the record was moved to the on-line "archive", where "semi-active" records were to be sent to wait out their retention periods, a procedure which was proposed to maintain fast response times on the most active records.

Terms of transfer

On the government side, there is often a delay of several years between the time records are appraised as archival and when they are transferred to the Archives. It has been our experience that there are at least two steps which can be taken to minimize problems at the time of transfer.

In Canada, a formal Agreement is signed with departments following appraisal, regardless of medium. It includes, as an appendix, a document outlining the Terms and Conditions of Transfer. These Terms and Conditions should be as specific as possible, defining the physical and logical formats for transfer. There should also be a description of what specifically was deemed archival in the application. Other clauses should address record extraction, format

conversion and the procedures relating to system upgrades while the application remains active.

The most important clause in the Terms and Conditions however is the "flexibility" clause. Technology and applications continues to evolve at a rapid pace and our Agreement with an institution must reflect that reality. For example, a department will probably have changed its physical storage format by the time the transfer occurs. It is likely that the Archives will also prefer a newer technology, but it is crucial that any change to the original Terms and Conditions be "mutually agreeable". Otherwise, the Archives could be forced to support every expensive, and often short-lived technology on the market.

The template document, which archivists use when drafting Terms and Conditions, has recently been updated by National Archives' archivists[1]. All electronic records are now categorized as one of 3 things:
1, structured records with linked digital objects
2. structured records without linked digital objects
3. individual digital objects

Monitoring

A second step in ensuring a successful transfer involves "monitoring". This concept was first introduced at the National Archives as we established our preference for acquiring electronic records at the end of their active life. This approach leaves systems in institutions for long periods of time, where they will undergo many upgrades and modifications. Monitoring helps ensure that what is available to the Archives at the end of a record's active life is what the archivist originally identified for retention.

Our current guideline states that any electronic system which has not been transferred to the Archives within two years of its appraisal must be "visited". This involves a review of the original appraisal report, the system documentation and the technical specifications, followed by a meeting with both the records manager and technical support personnel responsible for the records.

We prefer to meet both groups to overcome the large gap which still exists between records managers and information technology staff. In too many institutions, records managers do not understand automation, and technology staff are unaware of their obligations under the National Archives of Canada Act. We frequently find, for example, that the IT staff have thrown out their last tape drive for a certain physical carrier even though the Agreement with the Archives requires the transfer to take place on it.

Monitoring meetings allow us to identify system modifications which could create future problems for the Archives, to update our documentation, to learn about planned upgrades, and to remind the department of their commitment to the Archives to safeguard the records. In our first monitoring visits, we also realized that they provide an ideal opportunity to discuss larger records management issues in a very focused manner, and potentially to offer advice on the design of future systems.

Physical transfer

When it comes to the actual physical transfer, I will briefly list some of the problems we do run into on a regular basis:
- the tapes or diskettes received are blank
- they contain the wrong thing
- they are in an undecodable format
- they contain much more than we identified for transfer
- there is no documentation
- the documentation is inaccurate

---

[1] National Archives of Canada . *Proposed Procedures and Template for the Creation of Terms and Conditions for the Transfer of Archival Records,* Exposure draft, 25 April 2003. Terms and Conditions Work Group, Government Records Branch.

The first problem - the blank carriers - is the result of a human error. Somebody did not notice that the “copy” operation ended with an error message. The last one - inaccurate documentation - reflects the “live for today” attitude which is prevalent among those responsible for maintaining mission-critical operational systems. The middle four generally occur because of a breakdown in communication between the records management personnel, who are attempting to comply with the terms of their agreement with the National Archives, and the technical personnel who must make it happen.

We have not yet officially introduced electronic transfer, though there is an internal FTP site running which can be used to provide copies of holdings to the archivists. Though we are frequently asked about the possibility of electronic transfer, the demand is not yet significant. In most cases, problems with large file sizes, with privacy and security issues, and with firewalls at both ends of the transfer make it easier to move physical copies of the records. Also, with older systems, much of the documentation remains on paper and must be physically transferred anyway.

Processing

The obvious solution to the above problems is to process electronic records' immediately upon receipt, which is not often the case with paper records. You can then deal with the person in the department who actually prepared the records for transfer. The earlier a problem is discovered, the better the chance that you will be able to have the institution correct the problem and absorb the related costs. If the faulty transfer originated in a defunct institution, or an institution with little in-house expertise, there are only two options: de-accession the records or undertake the electronic equivalent of an archeological dig, trying to reconstruct the past from the surviving artifacts. The first creates an unnecessary gap in the historical record while the second is expensive and time consuming. If the records still exist in hard-copy, there may be a third option.

At the National Archives, the technical processing work is done by staff in a separate division - the Electronic Systems Project Division, recently renamed E-Operations. We continue to struggle with which functions are the archivist's job, and which are better done by the technician. Briefly, processing can include the following 15 steps:

1. counting the physical carriers and then reviewing and possibly listing the outer markings on them
2. reviewing any documentation provided
3. establishing an appropriate arrangement for the records, when this is lacking
4. creating a working copy
5. verifying the initial readability of the records
6. if applicable, comparing the data to the documentation, including record layout, field definitions and code values
7. possibly reducing duplication either within the electronic records, or between the paper and electronic components of an accession
8. creating a record of the previous step
9. assigning intellectual control numbers
10. creating archival descriptions
11. preparing and writing an introduction for the documentation package, either in hard-copy or electronic form
12. creating two copies for long-term preservation
13. assigning physical control numbers
14. establishing the need for a public use version of the data, or a specialized finding aid
15. defining requirements, if either of the above products is necessary.

This process is as time-consuming as it sounds. But the commitment to preserve electronic records, as with moving image and sound records, is an expensive one. The work must be done up front or later costs will escalate.

Description

As far as description is concerned, we upgraded the automated application which supports this function in the mid-1990s. The original system was implemented in 1984 and supported only file-level control in the tradition of the Anglo-American Cataloguing Rules' chapter on "computer files". In Canada, most archives now use the Rules for Archival Description to describe their holdings. Our current design is based on the draft of chapter 9 of RAD, which was developed in 1994. And I should add that the final, “official” version of that chapter was just released by the

Canadian Committee on Archival Description in January 2003. Our automated application allows hierarchical description, and contains the standard elements of archival description as well as a system description.

Preservation

Our preservation program has several facets. First, we maintain two copies of our electronic holdings. In the early years, both copies were on 9-track open-reel magnetic tape, but as the technology diversified in the mid-1980s, one copy was switched to optical disc (specifically 12" WORM), giving us additional protection against technological obsolescence.

In the early 1990s, with the imminent obsolescence of our optical disc format, we again studied the marketplace for possible replacements. An extensive study was completed in 1992 and recommended operational testing of 8mm EXABYTE cassettes and compact discs[2]. The 8mm is now in use but the compact disc format was never adopted because of its low storage capacity. Though I have not discussed them here today, the National Archives also acquires Geographic Information Systems, where large file sizes are the norm. 600 MB of storage per carrier was completely inadequate to accommodate their large file sizes. Instead, the conservators recommended Digital Linear Tape cartridges (DLT) for the second copy.

When 8mm tape was first introduced, many users criticized it, claiming that "read" problems developed rapidly - often within 12 to 18 months. Our experience with the format has been quite different. I asked one of the technical people why this might be. He explained that the manufacturer recommended "polishing" the tape before recording data onto it. This involved re-winding the cassette a couple of times, which we have consistently done before making preservation copies. It would appear that the lesson learned here is to read and follow the manufacturer's instructions.

It is now a generally accepted fact that the physical storage formats for electronic records will outlast their hardware and software, but it is still prudent to provide an appropriate storage environment. Until recently, our two copies were stored in separate locations. But a survey of some of our oldest magnetic tape formats was conducted in 2000. The results of the survey reinforced what conservators have been saying for years - stability in environmental storage conditions is key to media longevity. In our newest building - the Gatineau Preservation Centre which opened in 1997 - approximately 3% of the 9-track tapes were unreadable. Vault conditions in that building are extremely stable. In one of our older buildings, which was not purpose-built for archival storage - environmental conditions tend to fluctuate. There, almost 10% of the material tested exhibited some degree of "read" problem. As for the 8mm EXABYTE cassettes which, at the time of the survey, were on average 5 years old, no technical problems were detected at all. They did however turn up a "human" error rate of almost 2%, related to mis-labeled carriers and damage caused during cleaning.

Logical formats

Moving to logical storage formats, our preferred format for databases remains ASCII (more properly ISO 8859-1, the Latin 1 character set), though some of our earliest holdings are still in EBCDIC, a proprietary format used on IBM mainframes. We have not experienced any problems accessing EDBDIC files which were properly documented at the time of acquisition, but we have encountered problems with recent transfers of older material from outside sources. EBCDIC was extensible but we are finding that, 20 years after the fact, nobody in the creating organization can find any documentation about these "in-house" extensions. I often think of this when XML and its extensibility is discussed.

We have been accepting textual records in their native software format since 1989. Except for some conversion which was done to allow the initial acquisition, we have not yet faced the need to convert any of these holdings forward to maintain readability. But we know that we will not be able to count on our small collection of obsolete equipment and our selection of viewers to continue to provide access to the oldest documents for much longer. For the last two years, we have attempted to do some testing of conversion results from proprietary software forward to

[2] National Archives of Canada, *Report of the Working Group on Conservation Standards and Technologies*, 1992.

HTML or XML files. After converting a selection of documents automatically, we discovered the following problems:

- each conversion pass skipped at least one group of records. They were not copied. The gaps were different in the HTML and the XML versions.

- the first document we examined was a one-page agenda which had become a 3-page agenda having entirely lost the spacing of the original document.

- the second document examined randomly changed font sizes and types between, and even in the middle of paragraphs. There were no bad codes in the original document to explain this variance.

- in the final document that I looked at, all the footnotes had been lost.

This was not a scientific test conducted with absolute rigour. But with results like that, it would appear we have every reason to be concerned, first, about the reliability of currently available conversion software.

In recent years, we have also had to address the possible adoption of digital signature and encryption technology. In Canada, the system intended for adoption by the Canadian federal government is referred to as Public Key Infrastructure (or PKI for short) and is based on Entrust technology.

After having studied the technology, the National Archives issued Guidelines to federal government institutions. In short, these guidelines forbid the transfer of encrypted and digitally signed records to the National Archives, if the use of either has made the document unreadable without reliance on proprietary software technology - in other words, if using the current technology.

These guidelines suggest two further courses of action, which we have not yet pursued. First, the current Canadian record-keeping metadata standard should probably be revised to include optional data elements documenting the use of encryption and digital signatures, and recording successful un-encryption and verification of signatures. In addition, new records management advice addressing how to balance a department's immediate security concerns with our long-term archival concerns should probably be developed.

Access to Information and Privacy legislation

The review of electronic records in accordance with the Access to Information and Privacy Acts is still an infrequent event and is handled on a case-by-case basis. In one experiment, we have included several data elements to track the review of documents and their current access status in the electronic finding aid. This should allow researchers to move directly from the file list to the document itself, if it has already been reviewed and opened. We haven't been faced yet with the need to find the electronic equivalent of the black magic marker for textual records which have been partially opened. Volume remains so low that we continue to handle these situations in hard copy, preferring to focus instead on digitization projects for analog records.

Privacy issues are both more straightforward and more complicated when it comes to databases. It is, for the most part, a simple technical matter to remove personal and confidential data elements from a database when making a copy for a researcher.

The most difficult aspect of anonymization however is recognizing those data elements which appear innocuous, but which could be used to identify individuals. This requires a good understanding of the data and the frequency with which specific values and combinations of values appear in the database, as well as the possibility that various data matching schemes could re-identify the subject of an anonymized record. A good introduction to the issues can be found in a 1995 article published in the IASSIST Newsletter which outlines in some detail the anonymization of data samples from the 1991 British census prior to their public release[3].

[3] Dale, Angela. *Samples of Anonymised Records from the 1991 Census for Great Britain*, Newsletter - International Association for Social Science Information Service and Technology, summer 1995, pp. 5-12.

Reference

Reference requests against our electronic holdings remain low, despite the fact that they are far more visible in our revised, RAD-compliant descriptions than they were in the past in our “media-specific” systems. Given this situation, we are not currently devoting any resources to research and development in this area.

We believe we will be dealing with at least 3 quite different types of researchers. The traditional group is primarily interested in conducting statistical analyses of large data sets; they are technically literate and can usually draw on the extensive computer facilities and expertise in their home institution, often a university. We charge for labour costs for each file provided, as well as the cost of the physical carrier.

We continue to provide the traditional data reference service, first developed to provide datasets to researchers for statistical analysis. This involves creating a copy of a complete data file and its documentation and mailing it. We have also tested other delivery options, such as pre-cut CDs of popular holdings, as well as electronic distribution via Internet e-mail, FTP and download from a web site.

We have also had some experience with a second type of researcher, whose computer skills are limited to micro-computers and standard software packages. Reference services for this type of researcher often extend to helping them decipher system documentation, extracting sub-sets of data for them and converting the data to a more standard form, such as the .dbf file format so prevalent in commercial database software packages (except, of course, Microsoft’s). We will soon have to establish a policy and an appropriate fee structure for this type of "value-added" service, in situations where a researcher requests, or is entitled to see, only a portion of a data file.

A number of our holdings attract what we refer to as the "single record" user. These are researchers who are not interested in statistical analysis, but in individual records about specific people, or houses, or airplanes which, for whatever reason, are documented in databases which we hold. They generally have neither the skills nor the equipment needed to re-mount and search large data files. It is to provide appropriate services to this group that we first began experimenting with prototypes of public-use databases.

These reference products contain, first and foremost, unclassified data. They are also limited to that part of the database necessary to answer the most frequently asked questions. They have a simple user interface which guides the user through a limited number of search patterns and provides a formatted report for a single record.

We expect these products to sit on the National Archives web site, hyper-linked from the most detailed level of description prepared for the holding, which can be the series, file or item level, depending on the type of electronic record involved. The National Archives and Records Administration in Washington has developed this concept - the Access to Archival Databases service was recently made available on their web site.

The situation is somewhat different with textual records in electronic form. If the records cover a very narrow scope and exhibit some signs of arrangement, it is usually enough to explain the directory structure and document-naming protocol in the description.

For cases like the 350,000 electronic documents received from the Trade Negotiations Office however, we created a rather extensive item-level finding aid, which is a database in its own right and which will, some day, serve as the front-end to the records themselves. This approach was necessary to overcome the fact that the records were preserved from back-up tapes and were therefore organized technically, rather than intellectually, through the first three levels of directories. Specifically, the structure was based on the date of the backup, the server on which the user was assigned storage space for their files and finally, each user's first name, the naming protocol for which was changed at least three times during the short period the LAN was in use at TNO.

The finding aid we created equates each individual's various user names with a single standardized form of their name, as well as providing their position title and the section in which they worked. Secretaries to senior officials are specifically identified because, in this “automation period”, secretaries used computers but their bosses' generally didn’t. Additional data elements in the finding aid allow researchers to compare file titles, date and time stamps, and file sizes to locate draft and final versions of documents, and to track the movement of records across

time and through the directories of various employees.

Web sites

Just before concluding, I feel I must say a few words about web sites. Until the announcement of the merger of the National Archives and the National Library, we largely assumed web sites would be acquired and preserved by the National Library, a pattern which we have seen consistently on an international scale.

In 1999, we did prepare some advice which is available to government departments, but which was never formally issued by the Archives. Within an extensive discussion of web sites, accountability and risk assessment, the document suggests that records being placed on a web site should also be properly filed within an electronic records management system. As with the PKI situation, this suggests the possibility of some new metadata elements relating to the date a record was posted to a web site and the date it was removed and/or updated.

Having suggested that web sites were the National Library's responsibility, we did however acquire one single web site, perhaps in a moment of "mandate confusion". It can now be accessed, as an "archival holding", on our web site. And since the announcement of the merger, there have been some discussions with National Library staff who recently conducted a pilot project.

Conclusion

This has been a rather quick overview of the program at the National Archives of Canada.

Thank you